UNIVERSITY OF HAWAII
LIBRARY
JAN

The Philosophical Magazine

FIRST PUBLISHED IN 1798

A Journal of Theoretical Experimental and Applied Physics

Vol. 5 October 1960 No. 58
Eighth Series

25s. 0d., plus postage
Annual Subscription £13 10s. 0d., payable in advance

Printed and Published by
TAYLOR & FRANCIS LTD
RED LION COURT, FLEET STREET, LONDON, E.C.4

THE PHILOSOPHICAL MAGAZINE

Editor

Professor N. F. Mott, M.A., D.Sc., F.R.S.

Editorial Board

Sir Lawrence Bragg, O.B.E., M.C., M.A., D.Sc., F.R.S.
Sir George Thomson, M.A., D.Sc., F.R.S.
Professor A. M. Tyndall, C.B.E., D.Sc., F.R.S.

Authors wishing to submit papers for publication in the Journal should send manuscripts directly to the Publishers.

Manuscripts should be typed in *double* spacing on one side of quarto (8 × 10 in.) paper, and authors are urged to aim at absolute clarity of meaning and an attractive presentation of their texts.

References should be listed at the end in alphabetical order of authors and should be cited in the text in terms of author's name and date. Diagrams should normally be in Indian ink on white card, with lettering in soft pencil, the captions being typed on a separate sheet.

A leaflet giving detailed instructions to authors on the preparation of papers is available on request from the Publishers.

Authors are entitled to receive 25 offprints of a paper in the Journal free of charge, and additional offprints can be obtained from the Publishers.

The *Philosophical Magazine* and its companion journal, *Advances in Physics*, will accept papers for publication in experimental and theoretical physics. The *Philosophical Magazine* publishes contributions describing new results, letters to the editor and book reviews. *Advances in Physics* publishes articles surveying the present state of knowledge in any branch of the science in which recent progress has been made. The editors welcome contributions from overseas as well as from the United Kingdom, and papers may be published in English, French and German.

The Effects of Surface Condition on the Mechanical Properties of Lithium Fluoride Crystals†

By A. R. C. Westwood
RIAS, 7212 Bellona Avenue, Baltimore 12, Maryland, U.S.A.

[Received May 23, 1960]

Abstract

A comparative study has been made of the mechanical behaviour of as-cleaved, chemically polished and coated lithium fluoride crystals. Surface coatings were produced by vacuum deposition and by chemical reaction. Thin ($\gtrsim 10\,\mu$) reacted-coatings increased the yield stress some 50% and were responsible for catastrophic stress drops observed at the critical resolved shear stress. Coatings also decreased the initial rate of work-hardening and the stress and strain to fracture. Failure occurred by the formation of Stroh-type cracks.

An explanation for these and other results is offered based on considerations of the distribution and number of operable dislocation sources, and the piling up of dislocations at intersecting slip bands and beneath surface films. Metallographic evidence is presented for the latter phenomena. Experimental observations are more consistent with a barrier-mechanism of work-hardening than with a defect-mechanism.

§ 1. Introduction

Two concepts have proved useful in understanding some of the effects of surface films on the mechanical properties of metallic crystals. The first, due to Fisher (1952), is that a dislocation line emerging unpinned at the surface, but being pinned at some interior point, can serve as a source. Such sources are formed during crystal growth and should be activated at about half the stress required to activate interior sources of similar length. It follows that Fisher-type 'surface-sources' would be responsible for the initial stages of plastic deformation. However, a surface film might be expected to restrict or prevent the operation of these surface-sources, thus causing the increased yield stresses observed with coated metal crystals (Roscoe 1936, Adams and Cottrell 1958, and many other workers). Secondly, a surface film might prevent the egress of dislocations from a crystal so that they become piled-up beneath the film (Menter and Hall 1950, Phillips and Thompson 1950). Barrett's (1953) abnormal after-effect phenomenon can be explained in this way.

The objective of the present investigation was to obtain a better understanding of the effects of surface films on the mechanical properties of ionic crystals and to examine the validity of the above concepts for such crystals. For these studies, lithium fluoride was chosen in preference to other ionic

† Communicated by the Author.

crystals in order to take advantage of the dislocation-revealing techniques developed by Gilman and Johnston (1957).

Gilman *et al.* (1958) have demonstrated that a slowly moving cleavage crack (velocity $< 6 \times 10^{-3}$ cm/sec) can introduce many dislocation half-loops into the surface of lithium fluoride. These half-loops then serve as readily activatable sources during subsequent stressing. However, such artificially introduced surface sources, which must be distinguished from Fisher-type surface sources, can be completely removed by a short chemical polish.

To clarify the role played by these and other sources and by surface films in the yielding and work-hardening processes, comparative tests were performed on as-cleaved and chemically polished crystals in the coated and uncoated condition. In addition, some experiments aimed at providing direct evidence for the piling up of dislocations beneath a surface film were undertaken.

§ 2. Experimental Techniques

Lithium fluoride crystals were obtained from the Harshaw Chemical Company and compression specimens, approximately 12 mm–18 mm $\times$ 5 mm–6 mm square, were cleaved from these crystals. After the appropriate surface preparation, specimens were compressed in an Instron testing machine at strain rates of order 10^{-4} per sec.

For any particular test series, all specimens were cleaved from the same area of given crystal, were subjected to identical heat treatments and were subsequently handled with great care to prevent accidental surface damage. In this way, reasonable reproducibility of results was obtained ($\sim \pm 8\%$). Values of the yield stress, the critical resolved shear stress and the stress and strain to fracture were determined from the Instron recorder chart. The yield stress was taken as the stress at which stress–strain curves first deviated from linearity, and the critical resolved shear stress as the stress at which a large increase in strain occurred at almost constant stress. These two stresses are readily distinguished in the compressive stress–strain curves of lithium fluoride crystals.

Experiments were performed using specimens with a vacuum-deposited coating ($\gtrsim 5\,\mu$) of chromium or with a reacted-coating produced in the following way. Freshly cleaved or chemically polished crystals were gently placed in intimate contact with pure magnesium oxide powder. They were then slowly heated to 600°–700°c and held there for 1–3 hours. At these temperatures a chemical reaction producing MgF_2 and Li_2O occurs (Ludellans and Welsh 1952) and a coherent and slightly interdiffused surface coating results†. The specimens were then furnace-cooled slowly to 300°c

† The reacted-coating is believed to be an alloy layer, 5–50 μ thick, of magnesium in lithium fluoride. Magnesium fluoride is soluble in lithium fluoride (~ 6 mol. % max.) at temperatures above 400°c (Haven 1950) and can be precipitated out at lower temperatures. Elastic distortion in the lattice due to differences in ionic radii is negligible since Mg^{++} and Li^{+} are similar in size—0·78 Å radius. However, to preserve electrical neutrality, each magnesium ion is expected to be associated with a cation vacancy bound to it by the coulomb forces between them.

and finally air quenched to room temperature to avoid possible strain-ageing phenomena. Other specimens, not in contact with magnesium oxide, were given identical heat treatments for comparative tests.

For metallographic studies, the crystal preparation, polishing and etching techniques described by Gilman and Johnston (1957) and Johnston and Gilman (1959) were used.

§ 3. Experimental Observations

A thin film of chromium on as-grown, as-cleaved lithium fluoride crystals increased the yield stress some 12%, but no increase in critical resolved shear stress was observed and the subsequent work-hardening behaviour was not significantly affected. Examination of the chromium film after deformation revealed cracks which appeared to be associated with underlying bands of both edge and screw dislocations.

Reacted-coatings on as-cleaved specimens affected all stages of the deformation process. Table 1 displays the results of tests on specimens in the coated and uncoated conditions. The numerical values quoted are averages from tests on several specimens (3–6) in each condition and the figures in brackets are the percentage increase or decrease in value caused by the presence of the coating.

Table 1. The effect of reacted-coatings on the mechanical properties of as-cleaved lithium fluoride crystals. Change in property indicated in brackets

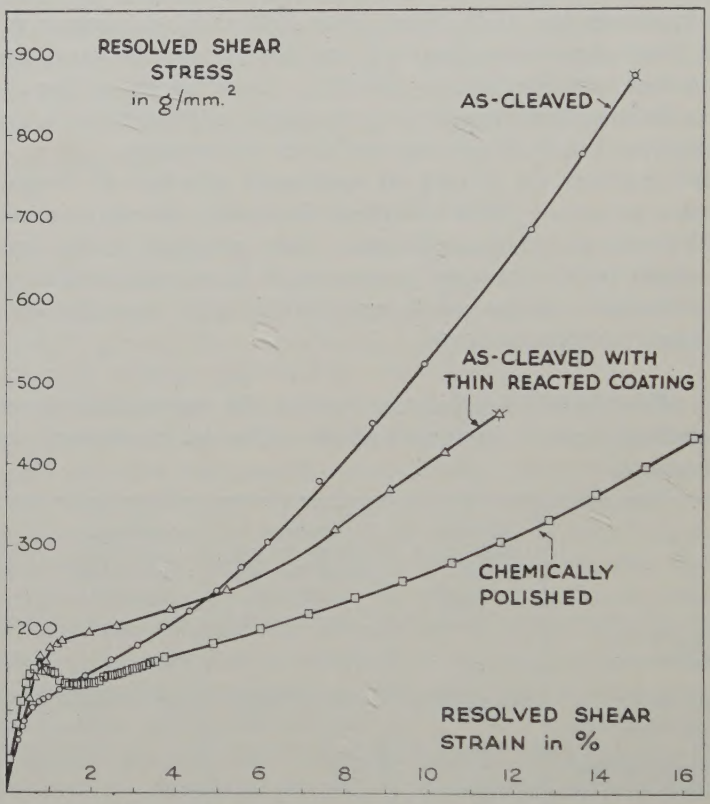
	Average yield stress in g/mm²	Average critical resolved shear stress in g/mm²	Average stress at fracture in g/mm²	Average glide strain at fracture in %
As-cleaved specimens from crystal block A	83	120	860	13·8
As above with thin ($\gtrsim 10\,\mu$) reacted-coating	128 (+54%)	149 (+25%), stress-drops	450 (−45%)	10·5 (−24%)
As-cleaved specimen from crystal block B	138	156	950	15·9
As above with thick ($25\,\mu$–$50\,\mu$) reacted-coating	222 (+61%)	265 (+70%)	505 (−47%)	8·0 (−50%)

Thin coatings ($\gtrsim 10\,\mu$) increased the yield stress by 50% and caused 'catastrophic' stress drops at the critical resolved shear stress. Thicker coatings ($\sim 25\,\mu$ and $\gtrsim 50\,\mu$) increased both the yield and critical resolved shear stresses by 60–70% but stress drops were not observed. However, the stress and strain to fracture were reduced to 50% of that of uncoated specimens. Figures 1 and 2† illustrate typical stress–strain curves and demonstrate that coated specimens exhibited initially a lower rate of work-hardening than similar uncoated crystals.

† With the exception of figs. 1, 2 and 8, the figures are shown as plates.

In contrast, a reacted-coating on a specimen chemically polished before the reaction treatment had very little effect on its yield stress (an increase 3%) and caused only a small increase in the critical resolved shear stress (13%). Table 2 summarizes the results of these tests and, for comparison, the results of others performed with chemically polished, annealed out of contact with magnesium oxide, and finally repolished specimens.

Fig. 1



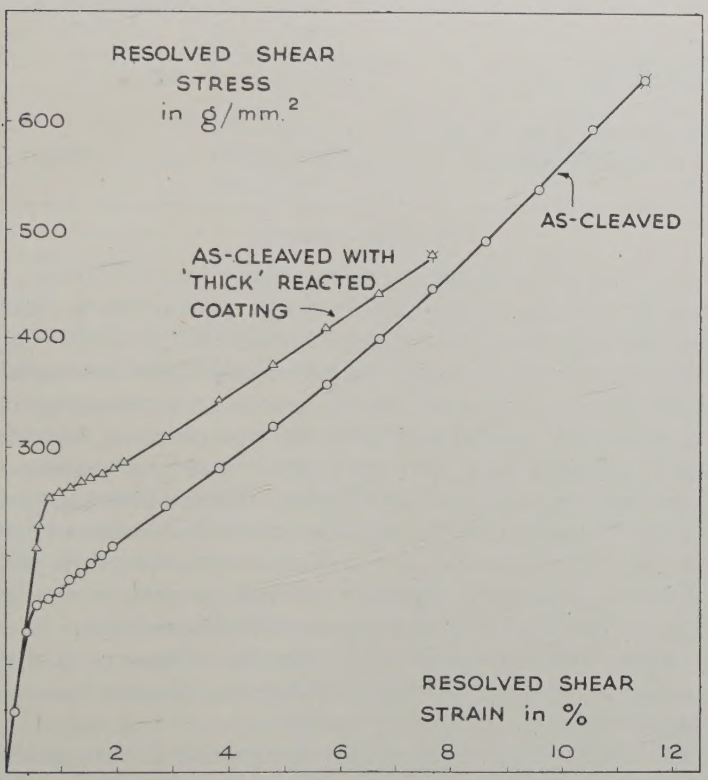
The effect of surface condition and thin reacted-coatings on the stress–strain curve of lithium fluoride crystals.

In all tests, failure was caused by nucleation of cracks (Stroh 1955) on (110) planes at slip band intersections—fig. 3. These cracks subsequently grew into slits and switched onto (100) cleavage planes in the manner described by Stokes *et al.* (1959) for magnesium oxide.

As an aid to understanding the above observations, metallographic studies were made on certain coated and deformed specimens. Evidence of dislocations piled up beneath the surface film and at other obstacles was found. Figure 4 shows edge dislocations piled up at a sub-boundary. It appears that such boundaries are not strong obstacles and usually cannot support more than 20–30 dislocations. The distribution of the

dislocations in fig. 4 does not conform to the relationships derived by Eshelby *et al.* (1951), probably because the dislocations relaxed back towards their source—not in the micrograph—between the straining and etching processes. Alternatively, Hirsch *et al.* (1959) have shown that, in stainless steel, piled-up groups of dislocations can interact with others on an

Fig. 2



The effect of a thicker reacted-coating on the stress–strain curve of an as-cleaved lithium fluoride crystal.

intersecting system to produce a network which can prevent the pile-up conforming to predicted relationships. In view of Johnston and Gilman's (1960) work on the form of dislocation loops in lithium fluoride, this alternative explanation is considered less likely.

Piled-up groups of screw dislocations are not common, but fig. 5 illustrates one example.

As in magnesium oxide, broad bands of edge dislocations act as barriers to edge dislocations on an intersecting slip system and are eventually responsible for the cracks which cause failure. Figure 6 shows a pile-up of edge dislocations that appear to have emanated from the half-loop source originally at A.

An example of a piled-up group of edge-dislocations beneath a reacted coating is given as fig. 7.

Table 2. The effect of 'reacted-coatings' on the yield behaviour of chemically polished lithium fluoride specimens from crystals block A

	Average yield stress in g/mm^2	Average critical resolved shear stress in g/mm^2
Chemically polished specimens	121	127
Chemically polished and reacted specimens	125 (+3%)	143 (+13%)

§ 4. Discussion

4.1. *The Effect of Surface Films on Yielding Behaviour*

Grown-in dislocations in lithium fluoride do not contribute significantly to the deformation process (Gilman and Johnston 1957) so that a comparison of the yield behaviour of as-cleaved and chemically polished crystals in the uncoated and coated conditions can be particularly revealing—table 3. Coating an as-cleaved crystal produces the same effect on the yield stress as a chemical polishing treatment, namely it causes an increase of some 50%. This can be explained as follows. Other investigations have demonstrated that small alloying additions of divalent cations to univalent cation type ionic crystals can exert a profound influence on the yield stress of such crystals, sometimes causing a tenfold increase in this property (Edner 1931, Schoenfield 1932, Nadeau and Washburn 1960). Hardening probably results from the increased difficulty of moving a dislocation through a lattice containing either foreign ion-vacancy pairs—due to mutual interactions—or precipitate particles.

In this investigation, reacted-coatings consisted of a coherent alloy layer of magnesium in lithium fluoride. No metallographic evidence for precipitated magnesium fluoride was obtained although its presence as submicroscopic particles is possible. Coatings affected the yield stress of the test specimens by preventing the operation of artificially introduced surface-sources. This could be achieved in two ways: by solute ion locking or by preventing half-loop expansion and multiplication by virtue of the increased lattice resistance to dislocation motion caused by the presence of magnesium ion-vacancy pairs or fine precipitate particles in the surface coating.

As-cleaved and coated specimens were thus forced to utilize other sources of dislocations also available to polished crystals, that is, interior sources. Gilman (1959) has suggested that such sources are dislocations introduced during cooling from heat-treatment temperature, by precipitate particles or other stress raisers and, in polycrystals, from grain boundaries. The small effect of a coating on the yield stress of chemically polished crystals suggests

that Fisher-type surface-sources—though possibly important in metal crystals—are not important in lithium fluoride crystals. This seems in accord with Gilman's (1959) studies which suggest that conventional Frank–Read sources are not responsible for dislocation multiplication in lithium fluoride.

Table 3. A comparison of the yield behaviour of as-cleaved and chemically polished crystals

	Uncoated		With 'thin' reacted-coating	
	Average yield stress in g/mm²	Average C.R.S.S. in g/mm²	Average yield stress in g/mm²	Average C.R.S.S. in g/mm²
As-cleaved crystals	83	120	128	149
Chemically polished crystals	121	125	127	143

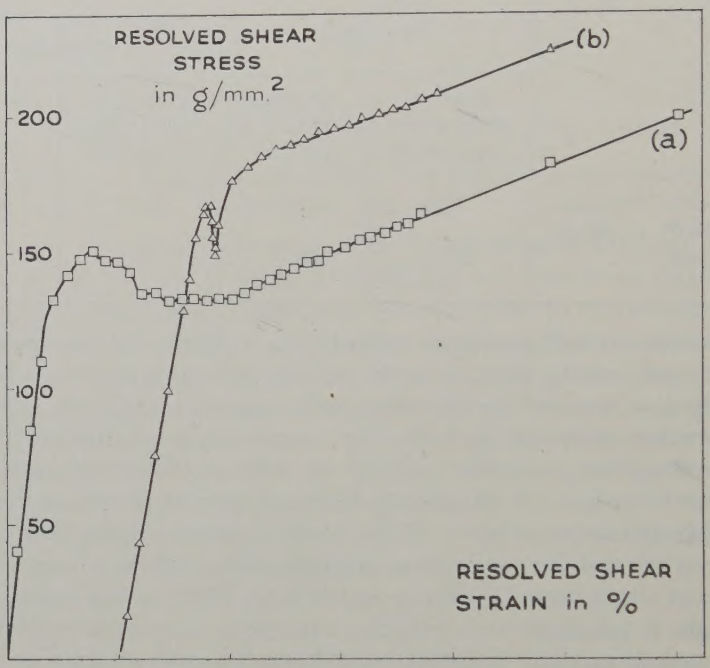
An as-cleaved and uncoated crystal then, deforms by the operation of many sources, mainly located at the surface and these begin to operate at relatively low stresses (in these crystals, approximately 80–100 g mm²). Multiplication proceeds probably by a cross-glide mechanism (Koehler 1952, Orowan 1954, Johnston and Gilman 1960) and at an increasing rate as the stress is raised. A chemically polished crystal, however, has fewer and relatively harder sources. Thus its yield stress is higher than a similar as-cleaved crystal, but since the multiplication rate is a very sensitive function of stress level (Johnston and Gilman 1959), when multiplication does begin it proceeds very rapidly. In some cases, the multiplication process in a chemically polished crystal reaches such a high rate that the specimen begins to deform at a greater rate than that caused by the motion of the testing machine crosshead and this is manifested by a meandering or 'casual' stress drop, fig. 8 (*a*).

The critical resolved shear stress is a more complex factor affected by such variables as strain rate, the number of dislocations present, their rate of multiplication and, in these studies, the ability of dislocations to pass out of the surface of the crystal.

Thin coatings of chromium had no significant effect on the critical resolved shear stress since they were themselves brittle and fractured during the compression test. Thus they did not constitute a real obstacle to the egress of dislocations. Relatively thin reacted-coatings however did provide an obstacle because of the increased resistance to dislocation motion in the alloyed layer and caused edge dislocations to pile up beneath the coatings —fig. 7. However, these piled-up groups were able to break through thin reacted coatings (5–10 μ) and cause a catastrophic drop in stress at the critical resolved shear stress, fig. 8 (*b*). Such catastrophic stress drops may be another manifestation of the high rate of multiplication of internal

sources once they are operated—compare the casual stress drop of polished crystals, fig. 8 (*a*). The production of a burst of dislocations which pile up and then break through a surface film appears to be a more likely cause of catastrophic stress drop than a process involving the steady piling up of dislocations until a break-through stress is reached. Stress drops were not observed with specimens having a thicker reacted coating, presumably because catastrophic break-through was not possible.

Fig. 8



Illustrating the ' casual ' and ' catastrophic ' stress-drops observed with lithium fluoride crystals.

4.2. *The Effect of Surface Coatings on Work-hardening and Fracture Behaviour*

During compression tests, four glide systems, ($\langle\bar{1}01\rangle/\{101\}$), were equally oriented to the applied stress, and glide on a single system was restricted to very low strains. However, the rate of work-hardening was linear over the first three to four per cent glide strain. Gilman and Johnston (1960) have also noted this and have suggested that work-hardening at low strains in lithium fluoride results from the production of lattice defects which are left in the wakes of moving dislocations. These defects then interfere with the movement of subsequent dislocations on the same or neighbouring slip planes. However, after 3 or 4% strain, the stress–strain curves, particularly of uncoated, as-cleaved crystals, deviate markedly from linearity—fig. 1. The present work suggests that a dislocation-barrier mechanism must be responsible for at least part of the work-hardening at

higher strains. The metallographic evidence of figure 6; a comparison of the work-hardening rates of as-cleaved and chemically polished crystals; observations on the effects of coatings on the stress–strain curve and the mode of fracture all appear consistent with this view. Based on this supposition, the experimental results can be explained as follows.

When an as-cleaved crystal is deformed, many sources operate and dislocations on intersecting systems soon meet and interact. Their mutual interference then results in a relatively high rate of work-hardening. Fewer sources operate in a polished crystal so that, at a similar strain, there should be more slip per source. This is evidenced by the presence of fewer, but broader, slip bands in such crystals. However, since the number of intersections is reduced, dislocation–dislocation interference effects should decrease so that a lower rate of work hardening should be, and is, observed with polished crystals; see fig. 1 and also Suzuki's (1957) studies with potassium chloride crystals.

A defect-hardening mechanism would predict the converse. More slip per source should result in greater defect-dislocation interference effects and a higher rate of work hardening for polished crystals than as-cleaved crystals. Thus the defect-mechanism does not satisfactorily explain work-hardening behaviour at higher strains.

By reducing the number of operable sources, a surface coating also reduces the initial rate of work-hardening. Specimens having a relatively thick coating (25–50 μ), probably deform by a leak out of dislocations through the film at a rate which is considerably less than the rate at which they pile-up beneath the film. In effect, the dislocations pile-up against a stable barrier so that the observed difference in fracture stress and strain between coated and uncoated specimens, table 1, may be accounted for as follows. A considerable strain is necessary to produce a sufficient number of dislocations concentrated into the thick bands which can act as a stable, rather than a temporary, barrier to edge dislocations on an intersecting system. However, for relatively thickly coated specimens, the coating itself can act as a suitable barrier. A crack nucleus then results either from the uncomplicated piling up of dislocations against this barrier or by the piling up of dislocations on a secondary system against a stabilized band of dislocations on a primary system—the stabilized band being caused by the piling up of dislocations beneath the surface coating. Observations on coated specimens showed that cracks always initiate at the surface and, apparently, just below the coating.

§ 5. Conclusions

The main conclusions from this investigation are:

(i) Metallographic evidence shows that surface coatings can act as barriers to the egress of edge-dislocations.

(ii) Tests on as-cleaved, chemically polished, coated and uncoated specimens, show that a coating can effectively restrict the operation of artificially introduced surface sources, thereby raising the yield stress. A coating also decreases the initial rate of work hardening of as-cleaved

lithium fluoride crystals by reducing the number of active sources so that mutual interference effects are lessened.

(iii) A catastrophic break through of a surface film by piled-up groups of dislocations can cause a sharp stress-drop. This can be readily distinguished from the more casual stress drop observed with uncoated chemically polished specimens but it is likely that both are caused by a high rate of dislocation multiplication.

(iv) A barrier-mechanism better explains work-hardening phenomena at strains above some three to four per cent.

(v) A surface coating can act as a stable barrier to dislocations causing pile-up, coalescence and the formation of crack nuclei beneath the film. This can reduce the stress and strain to fracture of coated crystals to half of that of uncoated crystals.

Acknowledgments

It is a pleasure to acknowledge the encouragement of Mr. W. W. Bender, the helpful discussions with Dr. H. M. Otte and the capable assistance of H. Opperhauser, Jr., D. Goldheim and C. Feltes in this work. We are also grateful for the financial support received from Watertown Arsenal Laboratories and in particular the continued interest of Mr. Murray Jacobson.

References

Adams, M. A., and Cottrell, A. H., 1958, *Acta. Met.*, **6,** 327.
Barrett, C. S., 1953, *Acta Met.*, **1,** 2.
Edner, A., 1931, *Z. Physik.*, **73,** 623.
Eshelby, J. D., Frank, F. C., and Nabarro, F. R. N., 1951, *Phil. Mag.*, **42,** 351.
Fisher, J. C., 1952, *Trans. Amer. Inst. min. (metall.) Engrs*, **194,** 531.
Gilman, J. J., 1959, *J. appl. Phys.*, **30,** 1584.
Gilman, J. J., Knudsen, C., and Walsh, W. P., 1958, *J. appl. Phys.*, **29,** 601.
Gilman, J. J., and Johnston, W. G., 1957, *Dislocations and the Mechanical Properties of Crystals* (New York : Wiley), p. 116 ; 1960, *J. appl. Phys.*, **31,** 687.
Haven, Y., 1950, *Rec. trav. Chim. Pays.-Bas.*, **69,** 1471.
Hirsch, P. B., Partridge, P. G., and Segall, R. L., 1959, *Phil. Mag.*, **4,** 721.
Johnston, W. G., and Gilman, J. J., 1959, *J. appl. Phys.*, **30,** 129 ; 1960, *Ibid.*, **31,** 632.
Koehler, J. S., 1952, *Phys. Rev.*, **86,** 52.
Ludellans, W. L. W., and Welsh, A. J. E., 1952, *Acta Cryst.*, **5,** 841.
Menter, J. M., and Hall, E. O., 1950, *Nature, Lond.*, **165,** 611.
Nadeau, J., and Washburn, J., 1960, 1st Tech. Report to U.S.A.E.C. Contract No. AT(11–1)–34. Project No. 28.
Orowan, E., 1954, *Dislocations in Metals* (New York : Amer. inst. min. (metall.) Engrs), p. 104.
Phillips, D. J., and Thompson, N., 1950, *Proc. phys. Soc. Lond.* B, **63,** 839.
Roscoe, R., 1936, *Phil. Mag.*, **21,** 399.
Schoenfield, H., 1932, *Z. Phys.*, **75,** 442.
Stokes, R. J., Johnston, T. L., and Li, C. H., 1959, *Phil. Mag.*, **4,** 920.
Stroh, A. N., 1955, *Proc. roy. Soc.* A, **232,** 548.
Suzuki, T., 1957, *Dislocations and the Mechanical Properties of Crystals* (New York : Wiley), p. 215.

Electron Microscope Observations of Deformed Magnesium Oxide†

By J. Washburn‡, G. W. Groves, A. Kelly and G. K. Williamson§

Department of Metallurgy, University of Cambridge

[Received May 9, 1960]

Abstract

Observations have been made by electron transmission microscopy of thin flakes of deformed magnesium oxide. The ease of cross slip of screw dislocations in this material has been directly verified. Cross slip leads to the formation of many markedly elongated prismatic loops of dislocation. These are observed profusely within slip bands, produced at room temperature. They are unstable at higher temperatures and break up into numbers of smaller prismatic loops. From these observations a mechanism for the widening of slip bands, observed under the optical microscope, is proposed. Certain dislocation interactions have been observed and analysed and preliminary observations made of slip band intersections.

§ 1. Introduction

The dislocation arrangement in deformed metal crystals has been extensively studied by electron transmission microscopy—see, for example, the review by Hirsch (1959). There have been no similar observations on non-metallic crystals. Etch pit techniques have been used to study dislocation motion and arrangement in deformed crystals of lithium fluoride and magnesium oxide (Gilman and Johnston 1956, Washburn *et al.* 1959, Stokes *et al.* 1959). However, very little information concerning the arrangement of dislocations within slip bands and at slip band intersections can be obtained because the density of etch pits becomes too great for optical resolution. A previous note (Washburn *et al.* 1960) reported that dislocations in magnesium oxide can be studied conveniently by electron transmission microscopy with an increase in resolving power of about two orders of magnitude.

The purpose of the present paper is to present some observations on the behaviour of dislocations in deformed magnesium oxide and to describe the dislocation networks within slip bands and at slip band intersections.

---

† Communicated by the Authors.

‡ National Science Foundation fellow on leave from the University of California.

§ On secondment from C.E.G.B. Berkeley Nuclear Laboratories.

§ 2. Experimental Procedure

Thin sheets of MgO, approximately $0{\cdot}025 \times 1 \times 2$ cm were cleaved on {100} faces from large crystals obtained from the Norton Company.

Specimens were chemically polished in 85% orthophosphoric acid at 100°C, to remove cleavage damage to the surfaces. After deforming plastically, specimens were etched for a few seconds in boiling concentrated nitric acid. The region selected for examination under the electron microscope was thinned by rotating the specimen slowly above a small jet of hot orthophosphoric acid. The specimen was removed from above the jet and washed in distilled water at the instant when a hole first appeared at some point. Clean surfaces were obtained by washing thoroughly in distilled water, and rinsing in methanol and then in ethyl ether. Areas sufficiently thin for electron transmission were revealed by the appearance of bright interference fringes in vertical optical illumination.

Observations were made with an electron microscope operating at either 80 or 100 kv. No trouble was encountered due to charging of specimens in the beam even though MgO is a poor conductor. It has been shown that the temperature rise of a metal specimen under the beam is $\leqslant 100$°C under normal operating conditions (Silcox and Whelan 1960). The temperature rise in MgO is likely to be considerably higher due to the lower thermal conductivity. Since application of the full beam current, by removing the condenser aperture, failed to evaporate a hole in the specimen if the area irradiated was close to one of the grid bars of the support it is felt that the temperature rise of the specimen under normal operating conditions was less than 800°C.

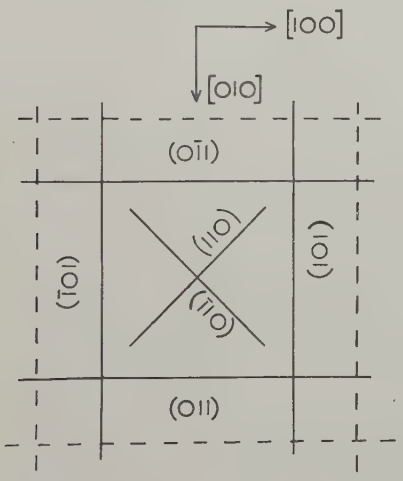
§ 3. Observations and Discussion

Most metal crystals studied in electron transmission have random crystallographic orientation with respect to the plane of the foil. In this work a ⟨100⟩ direction was always approximately parallel to the axis of the microscope, making interpretation of observations somewhat simpler. The shortest Burgers vector for a perfect dislocation in the MgO structure is $(a/2)$ ⟨110⟩. Etch pit studies have shown that this is the operative slip direction in MgO and that glide of dislocations takes place most easily on {110} planes. Interpretation of results is simpler than in metals for another reason: there is only one ⟨110⟩ slip direction in each {110} slip plane. Therefore a given slip plane is expected to contain only one kind of mobile dislocation. There are six {110} slip planes. If the surface of the foil is taken as (001), the projections of the {110} planes onto (001) appear as shown in fig. 1†. The four lying at 45° to the surface have a projected width equal to the thickness of the foil and intersect the surface along ⟨100⟩ directions. The remaining two are at 90° to the surface and project as lines lying along ⟨110⟩ directions. The

† All figures are shown as plates with the exception of figs. 1, 7, 8.

two different slip plane orientations relative to the foil normal will hereafter be referred to as 45° and 90° slip planes. The projection of a dislocation line lying in one of the 45° slip planes can have any direction and shape. Any dislocation in one of the 90° planes must appear either as a point or a straight line that lies exactly along ⟨110⟩. A dislocation in one of the 45° slip planes is in pure screw orientation when its projection lies along the ⟨100⟩ direction at right angles to the trace of its slip plane and in edge orientation when it lies along the trace of its slip plane. The ⟨110⟩ slip direction in the 90° planes is parallel to the surface. Therefore a dislocation in these planes is in screw orientation when it lies parallel to the foil surface and in pure edge orientation when it is at right angles to the surface of the foil.

Fig. 1



Projection of the {110} slip planes on to the (001) surface of an MgO flake.

3.1. *Motion of Screw Dislocations*

Figure 2 shows curved dislocations lying on one set of 45° planes. The diagonal line across the upper right-hand corner of the picture is a cleavage crack lying along ⟨100⟩. The dislocations had been moved by thermal stresses, generated by increasing the electron beam current, prior to taking the photograph. They appeared to be all of the same sign because they all moved towards the crack. The majority of these dislocations are approximately in screw orientation. Nearly straight screw dislocations were often found in the foils, but when they moved they frequently developed deep cusps, at points within the foil. These can be observed on some of the dislocations in fig. 2. The trails left on the surface by the ends of moving dislocations were not always parallel to the traces of the {110} slip planes.

Screw dislocations thus appear to be able to transfer easily from one {110} slip plane to a nearby parallel one. At the places where this transfer occurs jogs consisting of short segments of edge dislocation will be formed, which can only move by glide along the direction of the Burgers vector of the original screw. They will thus act as locking points and the screw will be bowed out between them. Figure 3 shows dislocations lying on 45° planes which were straight before local heating with the beam. They have all developed relatively immobile points and cusps along their length during motion.

3.2. *Dislocation Structure of Slip Bands*

The observations just described on moving dislocations were made in specimens containing a few isolated dislocations. To study the damage within slip bands crystals were bent at room temperature to produce slip on either 90° or 45° slip planes. 45° slip bands were obtained by bending about a ⟨100⟩ lying in the plane of the sheet, to an average surface strain of ~5%. To produce slip bands on 90° planes crystals were bent about an axis at right angles to the plane of the sheet.

45° slip bands appear as shown in fig. 4. Many dislocations can be observed terminating at top and bottom surfaces of the foil—this can often be recognized by a dark blob, at the end of the dislocation line. However, there are much larger numbers of dislocation lines which appear to be entirely within the foil and appear sometimes as a single line and sometimes as a double one. Since these dislocations do not appear to terminate at a surface or a node they must consist of a closed loop of dislocation. The majority of them run roughly parallel to a ⟨100⟩ direction but deviations from this direction of up to about 30° are quite common.

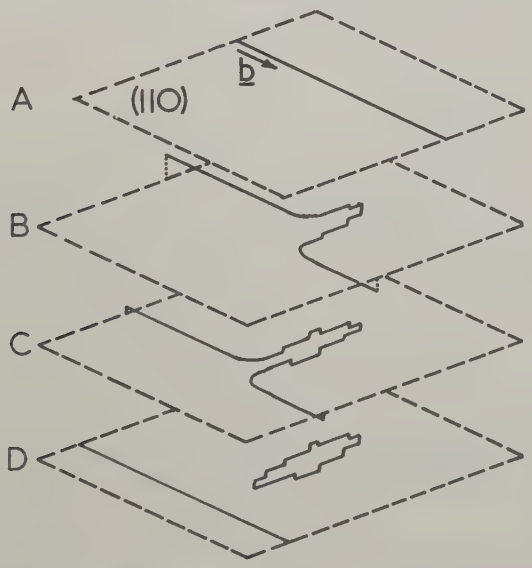
All dislocations in a given slip band had the same Burgers vector. This is expected from the geometry of the MgO structure and was verified in a number of cases by tilting the specimen into a position of zero diffraction contrast. All the dislocations in a band disappeared simultaneously at a setting when there was only one strongly diffracting plane parallel to the expected Burgers vector—an example is shown in fig. 5. The dislocation lines observed which do not terminate on the surface must therefore consist of a closed loop of dislocation which is predominantly of edge character. Since the majority of these are not resolved into separate dislocation lines the separation of the two sides of the loop must be less than a few hundred angstrom units. They will be designated, plus–minus pairs.

Figure 6 shows the dislocation population within a 90° slip band. The dislocations appear as lines approximately parallel to ⟨110⟩. Sometimes these lines appear double. Although these could be dislocations of predominantly screw character lying in a 90° plane it is thought more

likely, in view of the large numbers of plus–minus pairs observed in the 45° orientation of the slip plane that these are in fact also plus–minus pairs which deviate widely from a ⟨100⟩ direction. The fact that they sometimes appear as double lines, confirms this interpretation. These lines lie only approximately parallel to ⟨110⟩ which means that many dislocations do not lie exactly in the {110} slip planes, which again indicates cross slip.

The observations of plus–minus pairs and the fact that dislocations in slip bands do not lie exactly in a {110} plane are entirely consistent with the observations on moving screw dislocations and in fact the observations on the moving screws suggest the mode of genesis of the plus–minus pairs. This is illustrated in fig. 7. Lengths of screw dislocation are observed to cross slip out of one slip plane and then back to another parallel one. If

Fig. 7



Schematic illustration of the formation of plus–minus pairs. **b** is the Burgers vector of the dislocation.

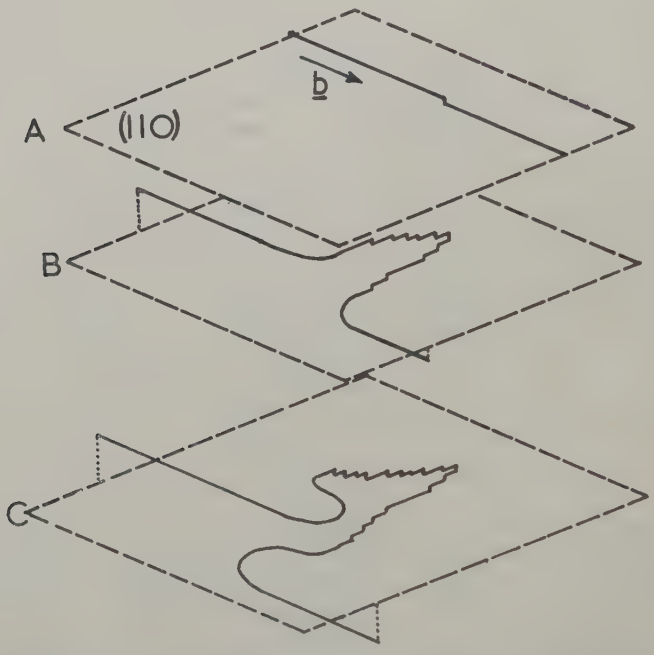
different parts of the same screw dislocation cross slip by different amounts then edge dislocation segments are formed and as the two parts of the screw continue to move a plus–minus pair of dislocations will be left behind the moving one. The plus–minus pair can be terminated if the two lengths of the screw (due to changes in the local stress field) come back to the same (110) plane again. The particular plus–minus pair will then become a closed prismatic loop and cannot be eliminated by glide alone; the locking point on the screw is then eliminated.

The dominant feature of slip bands observed under the electron microscope is the large number of plus–minus pairs. In view of this it is clear

that when etch pits are used to study dislocation distribution in slip bands many of the pits must mark the positions of close plus–minus pairs rather than those of single dislocations.

Dislocation multiplication may take place by a modification of the process suggested for the production of plus–minus pairs. During the formation of a pair the two lengths of screw dislocation may move apart by cross slip until they are sufficiently far apart for the force between them to be very small. Each could then move independently of the other forming two sources of dislocation on widely separated, but parallel slip planes. A schematic illustration of this process is given in fig. 8. It is proposed that such a process accounts for the formation of wide slip bands in MgO, observed under the optical microscope (Washburn *et al.* 1959), and for the increase in dislocation density when a single loop of dislocation is expanded in LiF crystals (Gilman and Johnston 1956).

Fig. 8



Schematic representation of a possible mechanism of dislocation multiplication.

3.3. *Stability of Plus–Minus Pairs*

The break-up of long plus–minus pairs into short lengths was observed in crystals deformed at a high temperature in an air–gas flame or could be observed directly when areas containing long pairs were heated in the microscope. An example of a pair breaking up during heating in the microscope by increasing the beam current is shown in fig. 9 (*a*) and (*b*). Since break-up is observed to have occurred in specimens which have not

been heated to greater than $\sim 1600°$C the process must occur in the absence of thermally generated point defects. If a pair of straight parallel edge dislocations of separation d breaks up into a number of circular prismatic loops each of radius R such that the area of prismatic loop projected normal to the Burgers vector is conserved, the ratio of the energy of the loops to the original plus–minus pair is given to a sufficient approximation by

$$\frac{d}{R}\frac{\ln R/r_0}{\ln d/r_0}$$

where r_0 is the radius of the core of a dislocation. The elastic energy of the system is then reduced if a plus–minus pair breaks up into a set of loops of average radius more than a few times greater than the original separation of the two dislocations in the plus–minus pair. This condition for stability reduces to the statement that the separation of the resultant loops should always be larger than 1·57 times their diameter. When the loops could be resolved this was found to be the case†.

3.4. *Dislocation Interactions*

In MgO each slip plane contains only one type of mobile dislocation, hence dislocation interactions are not expected to be observed as frequently as in face-centred cubic metals. They should occur predominantly at intersections of {110} planes. There are three possible reactions between the dislocations with Burgers vector $(a/2)$ ⟨110⟩ in this structure. Only one of these should lead to a reduction in elastic strain energy and it has been observed in a number of samples. A clear example of interactions of this type is shown in fig. 10. If the surface of the foil is taken as (001) dislocations A and B are on 45° slip planes at right angles to one another with Burgers vectors $(a/2)$ [011] and $(a/2)$ [0$\bar{1}$1] respectively. Both sets of dislocations have been intersected by dislocations C with Burgers vector $(a/2)$ [10$\bar{1}$].

The reaction between A and C gives

$$(a/2)\,[011]+(a/2)\,[10\bar{1}]\rightarrow(a/2)\,[110],$$

and between B and C,

$$(a/2)\,[0\bar{1}1]+(a/2)\,[10\bar{1}]\rightarrow(a/2)\,[1\bar{1}0].$$

The first reaction gives a short length of dislocation lying along [$\bar{1}$11] and the second one lying along [$\bar{1}\bar{1}$1]. These directions lie at 90° to one another in projection as is observed. In both cases the interaction produces a short length of dislocation which is pure edge and can only move in a {112} plane. The three dislocations meeting at any of the nodes do not have a common slip plane. This may be the reason why

† The break-up of long prismatic dislocations into a number of circular loops has recently been observed independently in zinc by P. B. Price (private communication).

the angles between the dislocation lines at the nodes are not equal to 120°. Alternatively this could be because a dislocation with Burgers vector ⟨110⟩ lying in {112} has a higher energy than one lying in {110}.

No other interactions have so far been unambiguously identified.

3.5. *Slip Band Intersections*

The intersections between slip bands are of particular interest from the point of view of crack nucleation. Cracks have been observed under the optical microscope to form at or near two types of slip band intersections. These are: (1) slip bands intersecting along ⟨100⟩, the two slip directions being at 90° to one another (Stokes *et al.* 1958, Washburn *et al.* 1959) and (2) slip bands intersecting along ⟨111⟩; the two slip directions being at 120° to one another (Keh *et al.* 1959).

The second type of intersection was easily studied under the electron microscope. It was produced between two 45° planes by bending a crystal about the [100] axis in the plane of the sheet and then straightening it and again bending about the [010]. An example of such an intersection is shown in fig. 11. It is characterized by a dense tangle of dislocations, many of which appear to be plus–minus pairs. It is not possible to follow the course of many dislocations at the intersection and no interactions can be definitely identified. There is no evidence that dislocations are piling up against the intersection. The increase in dislocation density appears confined to the intersection itself and the dislocation arrangement in the slip bands is the same at small and large distances from the intersection.

Intersections at right angles could be recognized at the intersection of two 90° slip bands—as illustrated in fig. 12. Nothing was observed at these intersections that looked like a pile-up of edge dislocations and no cracks were seen. However, examination of orthogonal intersections of slip bands in this orientation yields only limited information because the slip bands are viewed parallel to the slip plane and normal to the slip vector of the dislocations. It has not so far been possible to recognize unambiguously an orthogonal intersection of two 45° slip bands.

The electron microscope study of slip band intersections in these crystals has revealed that the intersection is characterized by an increase in density of plus–minus pairs but so far no cracks have been observed to form at them.

Acknowledgments

The authors are grateful to Professor A. H. Cottrell, F.R.S., for his encouragement throughout this research and to the National Science Foundation for the award of a fellowship to J.W. which made this cooperation possible. G.W.G. is grateful to the United Kingdom Atomic Energy Authority for the award of a bursary, and G.K.W. to the Central Electricity Generating Board for financial support.

References

Gilman J. J., and Johnston, W. G., 1956, *Dislocations and Mechanical Properties of Crystals* (New York: J. Wiley and Sons Inc.), p. 116.

Hirsch, P. B., 1959, *Metallurg. Rev.*, **4,** 101.

Keh, A. S., Li, J. C. M., and Chou, Y. T., 1959, *Acta Met.*, **7,** 694.

Silcox, J., and Whelan, M. J., 1960, *Phil. Mag.*, **5,** 1.

Stokes, R. J., Johnson, T. L., and Li, C. H., 1958, *Phil. Mag.*, **3,** 718; 1959, *Trans. Amer. Inst. min. (metall.) Engrs*, **215, 437**.

Washburn, J., Gorum, A. E., and Parker, E. R., 1959, *Trans. Amer. Inst. min. (metall.) Engrs*, **215,** 230.

Washburn, J., Kelly, A., and Williamson, G. K., 1960, *Phil. Mag.*, **5,** 192.

Cyclotron Resonance in Lead†

By J. E. Aubrey
The Royal Society Mond Laboratory, University of Cambridge‡

[Received May 11, 1960]

Abstract

Cyclotron resonance effects of the Azbel'–Kaner type have been observed in lead specimens at 9 kMc/sec and 1·2°K, and effective mass values of the charge carriers have been deduced from the magnetic field dependence of the surface resistance of the specimens. The anisotropy of the surface conductivity of the single crystal specimens in the low field limit was found to be consistent with the existing model of the Fermi surface in lead, and from the value of the surface conductivity of a polycrystalline specimen, a value of the total Fermi surface area was found which supports the view that the fourth Brillouin zone in lead is empty.

§ 1. Introduction

The theoretical prediction of Azbel' and Kaner (1956) that cyclotron resonance might be observable in metals when the magnetic field H is applied parallel to the sample surface, was verified by Aubrey and Chambers (1957) working on bismuth, and on completion of some additional experiments on bismuth (Aubrey 1960), it was decided to make a preliminary study of lead§.

Azbel' and Kaner showed that for the simple case of a quadratic electron energy band, the surface impedance Z of the metal in the anomalous skin effect limit varies with magnetic field according to the relation

$$Z \simeq Z(0)[1-\exp(-2\pi i x)\exp(-2\pi x/\omega\tau)]^{1/3} \quad . \quad . \quad . \quad (1)$$

where $Z(0)$ is the impedance in zero field and $x=\omega/\omega_c=m^*c\omega/eH$. For $\omega\tau \gg 1$, both real and imaginary parts of Z are oscillatory functions of x with period $\Delta x=1$, and from a measurement of the period of the oscillations in H^{-1}, the effective mass m^* of the electrons can be found. In the work to be described, the field dependence of the real part, R, of Z has been investigated in the range of magnetic field 1200 gauss‖ to 9000 gauss by

† Communicated by the Author.

‡ Present address: Department of Physics, University of Pennsylvania, Philadelphia 4, Pennsylvania, U.S.A.

§ Bezuglyi and Galkin (1958) have also done some preliminary work on lead using a very similar experimental method to that used in the present work but have published results on only one sample whose orientation was not specified.

‖ The field 1200 gauss was approximately the lowest field at which the transition from the superconducting to the normal state was complete.

observing the change in the Q of a resonator of which the specimen forms the most dissipative part: the specimens were wires, 1 mm in diameter and 1·4 cm long, constituting the central conductor of a coaxial resonator, whose design has been described by Pippard (1950).

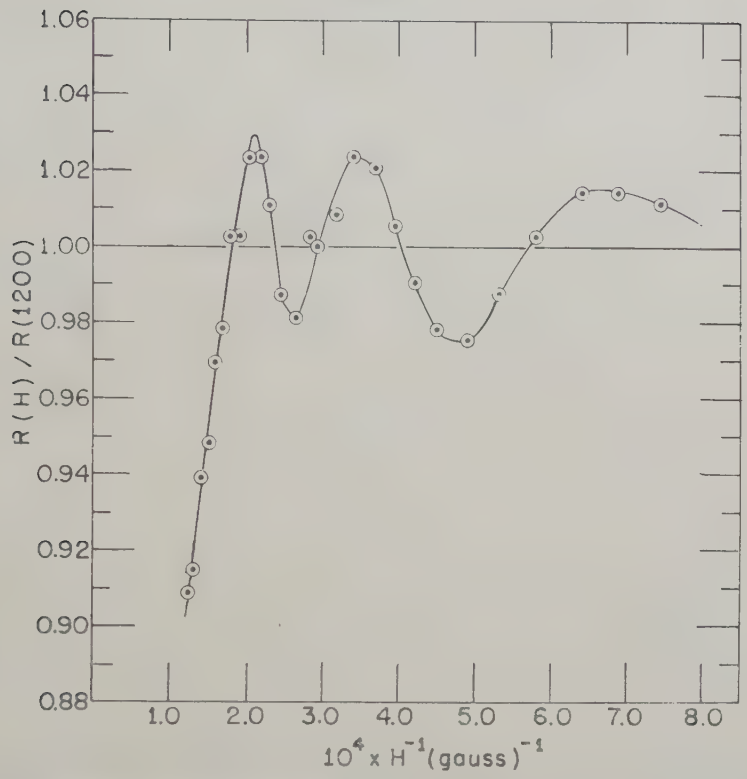
Zone-refined lead with resistance ratio $\rho(290°\text{K})/\rho(1{\cdot}9°\text{K}) = 2{\cdot}5 \times 10^4$ supplied by the Westinghouse Research Laboratories was used, and to form wire for the specimens, the lead billets were first extruded at room temperature through a 1 mm steel die. The extruded metal was etched with concentrated hydrochloric acid to remove contamination picked up from the die, and single crystals were grown on a flat plate furnace. These were subsequently electropolished using the perchloric and acetic acid mixture specified by Tegart (1956) yielding a highly specular surface whose brightness did not diminish between polishing and insertion of the specimen into the cryostat. Some care was needed in mounting the specimens to ensure that the magnetic field, provided by an air-cored solenoid, was as nearly parallel to the specimen axis as possible.

§ 2. Results

Specimens whose axes had the crystallographic orientations (100), (110), (111), (210), (211), (221), were studied and curves of $R(H)/R(1200)$ versus H^{-1} for each of these orientations are given in figs. 1 to 6. Each of the curves shows maxima and minima whose positions were found to be reproducible in repetitions of the measurements under identical conditions, and the values of H^{-1} at which these occur are given in the table. It is clear that in general there is no precise periodicity in H^{-1} of the kind predicted by eqn. (1), except in fig. 4 and possibly fig. 5. Thus it is clear that the observed behaviour cannot generally be accounted for by the presence in the metal of a single group of charge carriers of unique effective mass.

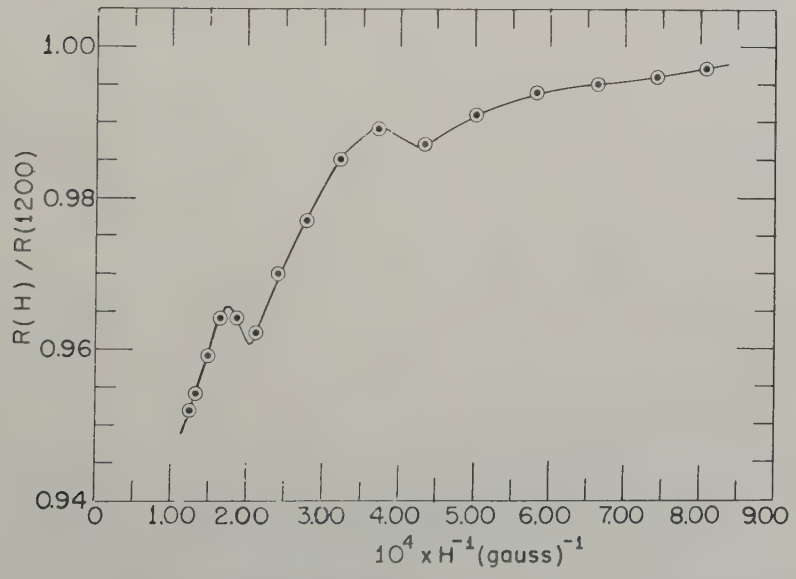
Orientation	Resistance maxima $10^4 \times H^{-1}$ (gauss)$^{-1}$	Resistance minima $10^4 \times H^{-1}$ (gauss)$^{-1}$	m^*/m_0
100	2·1, 3·5, 6·6	2·6, 4·8	$1{\cdot}3,\ 0{\cdot}6_6$
110	1·8, 3·7	2·1, 4·2	$1{\cdot}6,\ 0{\cdot}7_6$
111	2·1, 3·6	2·4, 4·3	$1{\cdot}4,\ 0{\cdot}7_4$
210	1·7, 3·5 5·6, 7·7	2·0, 4·4 6·5, 8·4	$1{\cdot}7,\ 0{\cdot}7_2$ or $1{\cdot}4_4 \pm 0{\cdot}0_7$
211	1·8, 3·5 5·9, 8·4	2·3, 4·8 6·6	$1{\cdot}4,\ 0{\cdot}6_6$ or $1{\cdot}3_7 \pm 0{\cdot}1_9$
221	1·4, 2·0, 3·1 4·4, 6·5, 8·5	1·7, 2·6, 3·8 5·3, 7·3	2·0, 1·3

Fig. 1



Surface resistance as a function of reciprocal field for a (100) specimen.

Fig. 2



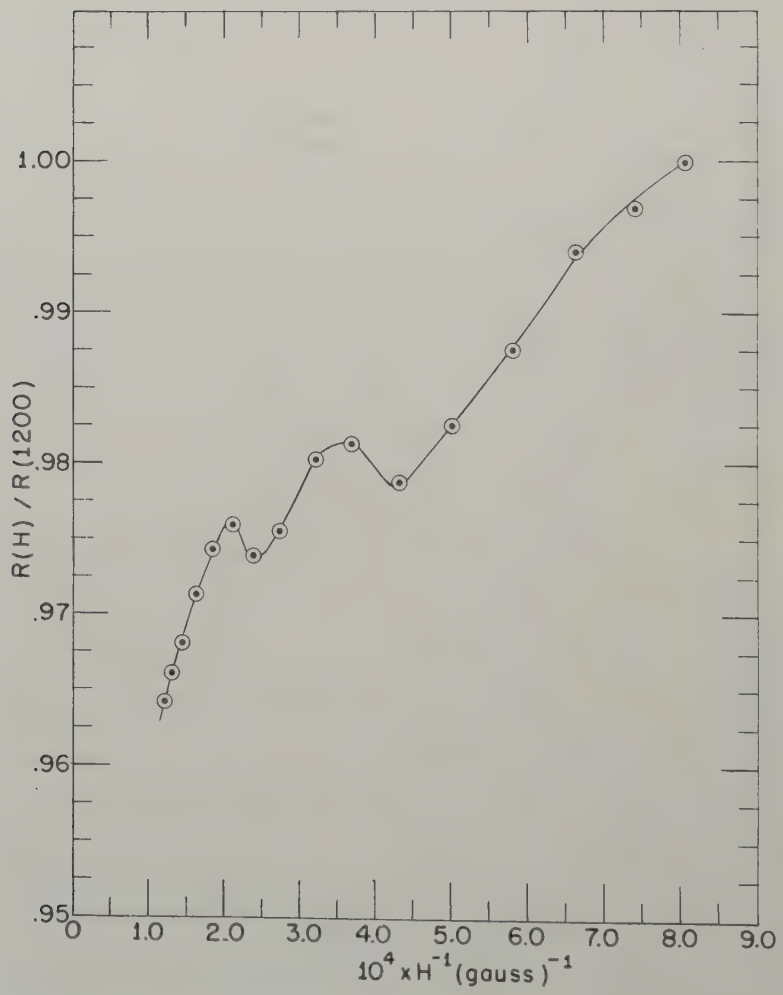
Surface resistance as a function of reciprocal field for a (110) specimen.

For a multiple band model, the form of the theoretical curve of surface resistance against reciprocal field may be most simply obtained from the approach used by Heine (1957) though the results follow also from the original form of the theory as given by Azbel' and Kaner (1957). For the present purpose it is sufficient to consider the simple case of a band structure in which the Fermi surface consists of a number of spheres, and we obtain from Heine's eqns. (6) and (12):

$$Z = 2\exp(i\pi/3)\,(2\pi^2\omega^2/s)^{1/3}, \quad s = \sum_{\nu} a_{\nu} \qquad . \; . \; . \; . \; . \qquad (2)$$

where a_{ν} involves only the band indexed ν, and the summation is over the total number of bands.

Fig. 3

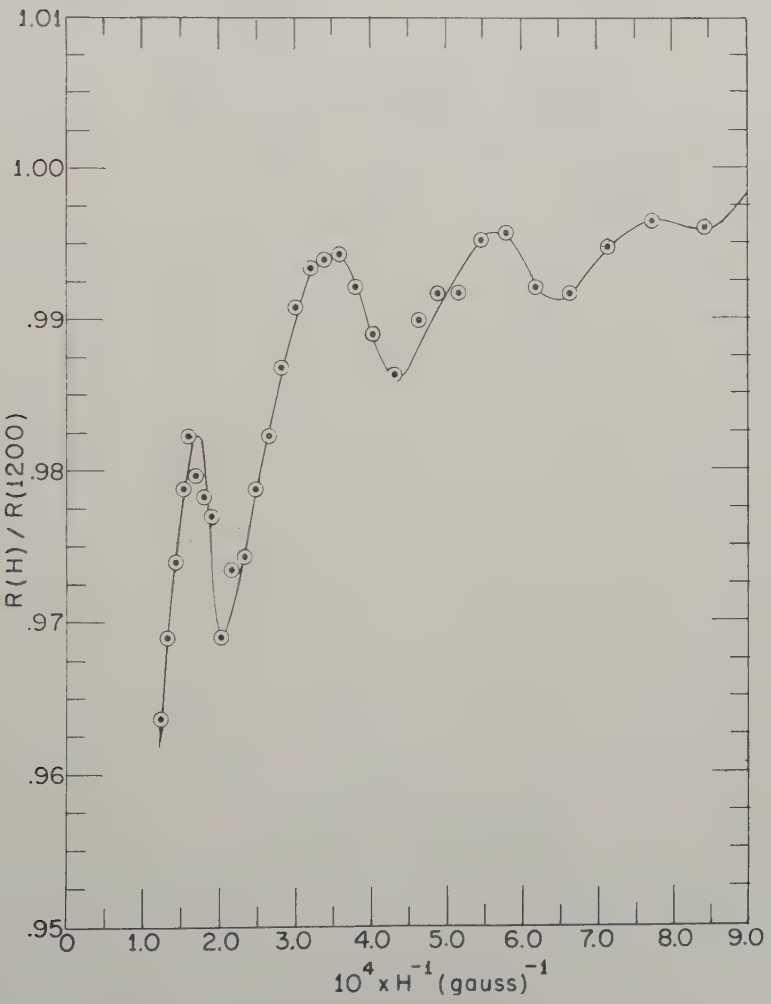


Surface resistance as a function of reciprocal field for a (111) specimen

Figure 7 shows the variation of $R(H)/R(0)$ with H^{-1} as calculated from (2) for the special case of a two band model with effective masses in the ratio 2 : 1, taking $\omega\tau = 2\pi$ in each band, and assuming that both bands contribute equally to the effective conductivity in zero field. The number of bands is restricted here to two partly for the sake of simplicity, and partly because the two-band curve reproduces the behaviour observed in figs. 1 to 3 fairly adequately. Included in fig. 7 are curves of $R(H)/R(0)$ for the two constituent bands of the model considered separately (eqn. (1)). The fundamental resonance (i.e. the first resistance minimum), which occurs at approximately

$$x = 1{\cdot}1 = \frac{m^* c\omega}{eH_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

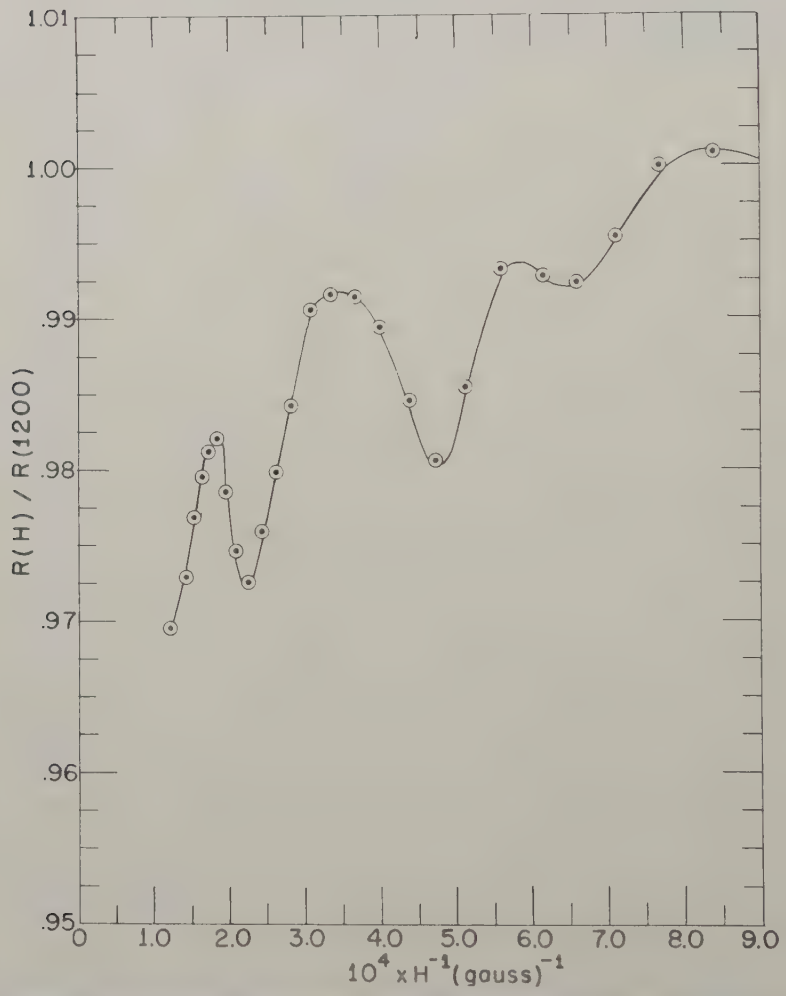
Fig. 4



Surface resistance as a function of reciprocal field for a (210) specimen.

for each individual band alone, occurs at virtually the same position in the composite pattern†. This result is rather insensitive to the precise ratio of the effective masses used in the model and also to changes in the assumed values of $\omega\tau$ in the two bands and in their relative contributions to the effective conductivity. We shall therefore use (3) in interpreting the experimental results, i.e. we apply (3) to the positions H_1, of the first

Fig. 5

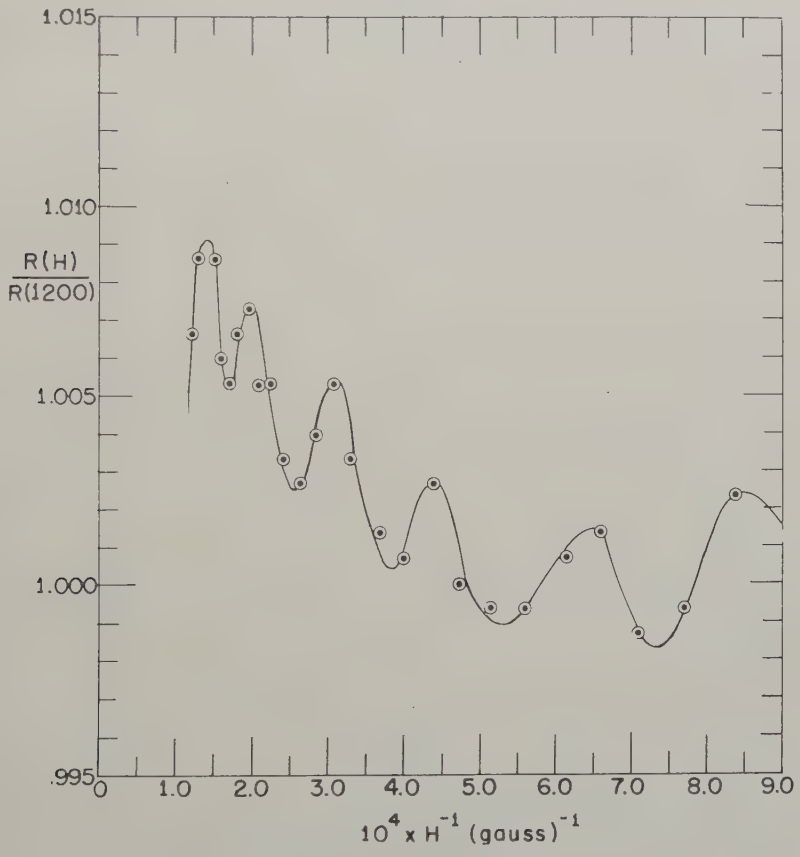


Surface resistance as a function of reciprocal field for a (211) specimen.

† The minima in R are related directly to maxima in the a_ν (eqn. (2)) which are themselves periodic in H^{-1} for each band. The fundamental maxima in the a_ν are sufficiently strong, with a realistic choice of $\omega\tau$, for the corresponding minima in the two-band resistance curve to appear at virtually unchanged positions with reference to the single band curves.

two resistance minima as given in the table and thus obtain the values of m^* shown in the fourth column of the table. The error in these mass values arising from the uncertainty of about $0{\cdot}2 \times 10^{-4}$ (gauss)$^{-1}$ in the positions of the resistance minima amounts to 10% for the set of larger mass values and 5% for the smaller values. A more important source of error however lies in the possibility that the phase of the oscillations in the surface resistance is not as predicted by the Azbel'–Kaner theory. Phase shifts have in fact been found in the work on bismuth (Aubrey and Chambers 1957) and also in work on zinc (Galt *et al.* 1959) and on copper (Langenberg and Moore 1959), and when present they may lead to errors, possibly of as much as 20 or 30% in the mass values calculated using eqn. (3).

Fig. 6



Surface resistance as a function of reciprocal field for a (221) specimen.

An alternative interpretation seems plausible for figs. 4 and 5, namely, that each of the curves reflects primarily the presence of a group of carriers with a single effective mass. The mean periods, $\Delta(H^{-1})$, are found to be

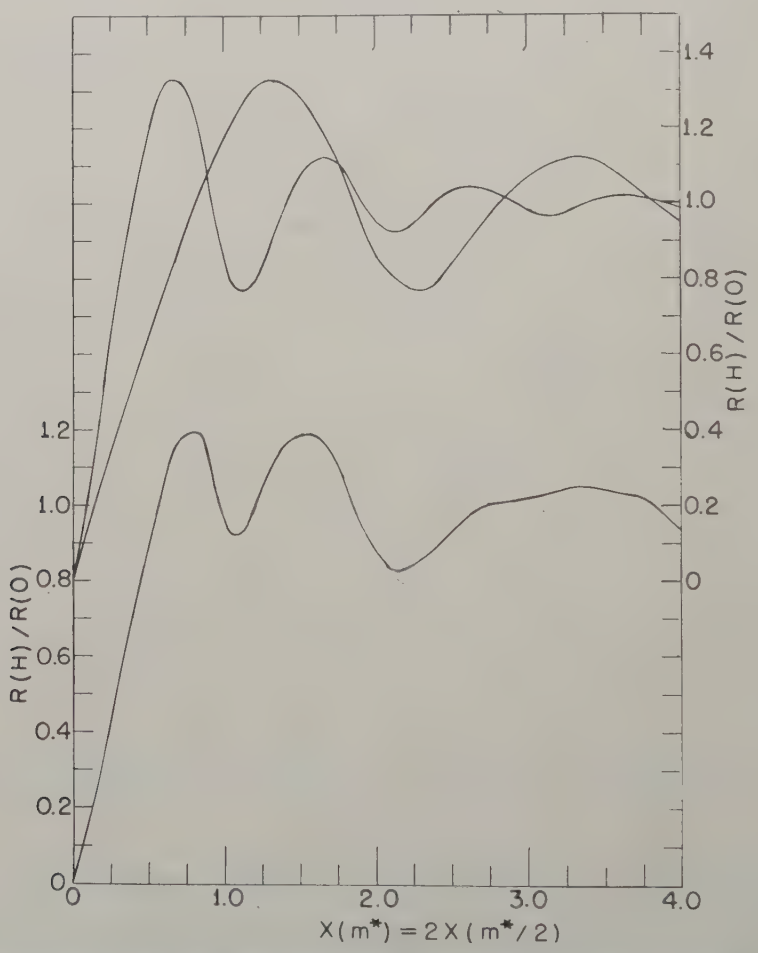
$$2{\cdot}1 \pm 0{\cdot}1 \times 10^{-4} \quad \text{and} \quad 2{\cdot}2 \pm 0{\cdot}3 \times 10^{-4}\,(\text{gauss})^{-1},$$

and the corresponding effective mass values calculated from the relation

$$m^* = \frac{e}{\omega c \Delta(H^{-1})} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

are recorded as alternative values in the fourth column of the table. The fact that the value for the (210) orientation differs appreciably from the value calculated from the position of the first resistance minimum means that the curve, treated as a single Azbel'–Kaner oscillatory pattern, has a phase shift of the kind mentioned above, and confirms the possibility that there may be considerable errors in the effective mass values calculated from the positions of resistance minima using eqn. (3).

Fig. 7



Upper curves (right-hand ordinate scale): theoretical surface resistance curves for the single band model (eqn. (1)). The abscissa

$$x(m^*) = m^* c\omega / eH = 2\,x(m^*/2).$$

Lower curve (left-hand ordinate scale): theoretical surface resistance curve for the two band model described in the text.

§ 3. Interpretation of the Results

Gold (1958) has suggested a model of the Fermi surface of lead from an analysis of the de Haas–van Alphen effect, using the free electron model as a guide. According to this model, electrons occupy the first four Brillouin zones in k-space, the first being completely filled, the second having a vacant central region, the third being occupied along the zone edges and the fourth being occupied at the zone corners. Recent considerations (Cohen and Gold 1960) have led to the suggestion that the fourth zone is in fact empty, so that portions of the Fermi surface exist in the second and third zones only. Gold finds an approximately isotropic effective mass of the order $1{\cdot}0 \pm 0{\cdot}1 m_0$ for the holes of the second zone, and a value $0{\cdot}56 \pm 0{\cdot}06 m_0$ for the electrons moving in orbits about the 'limbs' of the multiply connected surface in the third zone; it seems plausible therefore to associate the larger set of mass values, which fall in the range 1·3 to $2{\cdot}0 m_0$ with the second zone holes, and the set of values in the range $0{\cdot}6_6$ to $0{\cdot}7_6 m_0$ with the third zone electrons. Exact agreement between Gold's values and those found here is not to be expected since the de Haas–van Alphen values refer to electrons moving in orbits of extremal area on the Fermi surface, whereas the cyclotron resonance mass values refer to electrons moving in orbits on regions of the Fermi surface where the effective mass itself passes through extremal values (Azbel' and Kaner 1957).

§ 4. R.F. Conductivities

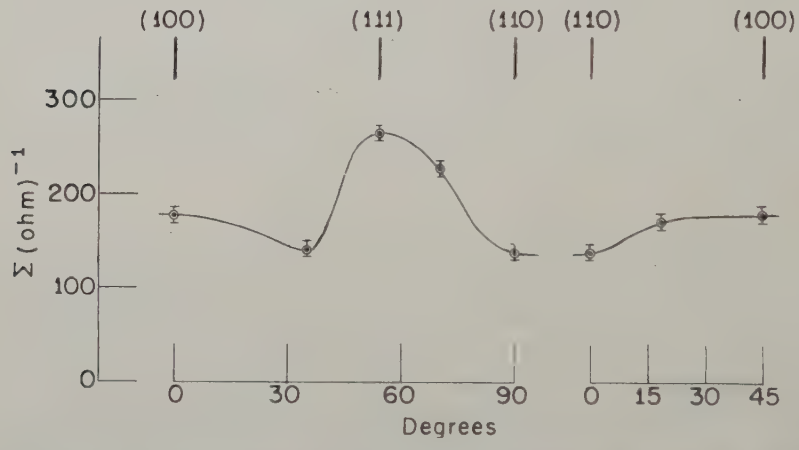
Useful information about the band structure can be obtained from measurements of the absolute value of the surface conductivity, Σ, at the anomalous skin effect limit in zero magnetic field†. Thus if plane single crystal specimens of a sufficient number of orientations are measured, the complete shape of the Fermi surface can in principle be found. This method is feasible in practice only if the Fermi suface has a relatively simple shape, e.g. if it consists of a single sheet, only, as in copper (Pippard 1957), or if the individual sheets have a simple analytic form, as in bismuth (Smith 1959). In the present work the measured value of Σ is an average over the azimuth of the cylindrical-shaped specimens, and all that can be attempted is a qualitative test of the compatibility of the observed anisotropy of Σ (fig. 8) with the existing model of the Fermi surface.

For a plane single crystal specimen (or at any azimuth of a cylindrical-shaped single crystal specimen), the principal components of the surface conductivity tensor are determined by integrals over the Fermi surface of the form, $\int |\rho(k_x)|\, dk_x$ and $\int |\rho(k_y)|\, dk_y$ (Pippard 1954). Here k_x and k_y are chosen as coordinates in k-space such that the corresponding axes in real space, x and y, are parallel to the principal directions of the surface impedance tensor. The quantity $|\rho(k_x)|$ is the absolute magnitude of the

† At temperatures below 7°K this could not be observed directly because the lead specimens became superconducting in fields below about 1200 gauss. Extrapolation from higher fields, however, involved little error: see figs. 1 to 6.

radius of curvature of the section of the Fermi surface in the plane k_x at the point where the normal to the section is parallel to k_y, and similarly for $|\rho(k_y)|$. Thus if $|\rho(k_x)|$ for example is large, the section dk_x about k_x makes a large contribution to the surface current, a large total surface current indicating a high value of Σ. It would seem reasonable to expect an almost isotropic contribution to the surface current from the holes of the second zone on account of the isotropy of the de Haas–van Alphen period and the distance of the Fermi surface from the Brillouin zone boundary (the surface contains approximately 0·4 holes per atom). Most of the observed anisotropy must therefore be due to the third zone electrons which occupy the multiply connected surface illustrated schematically in fig. 7 of Gold (1958). The limbs of this surface were shown as cylinders, but a more satisfactory fit of the period data (the γ oscillations in Gold's

Fig. 8



Anisotropy of the surface conductivity of the single crystal specimens.

paper) can be obtained by taking a prolate spheroid with axial ratio 2·3 : 1 as the model of one of the limbs. If a Fermi surface consisting of a single prolate spheroid is considered, the surface current will be largest when the direction of current flow makes an angle of 90° with the spheroid axis, for in this case the radii of curvature of sections of the spheroid are large. This implies that the spheroid axis is normal to the metal surface, a condition which can hold only if the specimen surface is plane. For a cylindrical specimen with the spheroid axis normal to the specimen axis, the surface conductivity will vary over the azimuth, and there will be two diametrically opposite regions where Σ is large. Since six of the twelve spheroids which make up the repeat unit of the multiply connected surface have their axes normal to the direction of current flow for a (111) specimen we should expect a maximum in Σ for this orientation. Four spheroids

satisfy this condition for a (100) specimen so we should again expect the surface conductivity to be large at this orientation, though not as large as at (111). These brief considerations explain the qualitative features of fig. 8, and it is likely that a model of the Fermi surface made up of twelve interconnected spheroids, and a single sphere representing holes of the second zone, would yield the observed anisotropy of Σ.

Finally, the total area of the Fermi surface, S, and the related parameter σ/l where σ is the d.c. conductivity, and l is an average value of the mean free path, can be calculated from the surface conductivity of a polycrystalline specimen. The appropriate relations are

$$S = \frac{2\pi^3 3^{3/2} hc^2 \omega^2 \Sigma^3}{e^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

(Chambers 1952), and

$$\sigma/l = \sqrt{3}\pi\omega^2\Sigma^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

(Reuter and Sondheimer 1948). Σ for a polycrystalline specimen was found to be $181 \pm 6\,(\text{ohm})^{-1}$ and this leads to the values

$$S = 17{\cdot}1 \pm 1{\cdot}7 \times 10^{16}\,\text{cm}^{-2},$$

$$\sigma/l = 11{\cdot}1 \pm 1{\cdot}1 \times 10^{10}\,(\text{ohm cm}^2)^{-1}.$$

These are in reasonably good agreement with the value of σ/l obtained from measurements at 1·2 and 3·6 kMc/sec by Chambers (1952), namely $9{\cdot}4 \pm 0{\cdot}7 \times 10^{10}\,(\text{ohm cm}^2)^{-1}$; the somewhat higher values found here can be accounted for by the better surface finish obtained in this work. Gold's model of the Fermi surface leads to a value for S of $25{\cdot}5 \times 10^{16}\,\text{cm}^{-2}$ if the corners of the fourth zone are occupied, or $18{\cdot}6 \times 10^{16}\,\text{cm}^{-2}$ if they are unoccupied. The experimental value given above thus supports the belief that the fourth zone is empty.

Acknowledgments

I wish to thank Dr. D. Shoenberg, F.R.S., and Dr. R. G. Chambers for their interest in this work and for their assistance in the preparation of the manuscript, and Dr. A. V. Gold for several stimulating discussions. I wish also to express my gratitude to the Department of Scientific and Industrial Research for the award of a maintenance grant, and to the U.S. Atomic Energy Commission for support on contract number AT(30–1)2395 during the period of writing.

References

Aubrey, J. E., 1960, *J. Phys. Chem Solids* (to be published).
Aubrey, J. E., and Chambers, R. G., 1957, *J. Phys. Chem. Solids*, **3**, 128.
Azbel', M. Ya., and Kaner, E. A., 1956, *J. exp. theor. Phys., Moscow*, **3**, 772; 1957, *Ibid.*, **5**, 730.
Bezuglyi, P., and Galkin, A., 1958, *J. exp. theor. Phys., Moscow*, **7**, 163.
Chambers, R. G., 1952, *Proc. roy. Soc.* A, **215**, 481.
Cohen, M. H., and Gold, A. V., 1960, *Phil. Mag.* (to be published).

Galt, J. K., Merritt, F. R., Yager, W. A., and Dail, H. W., Jr., 1959, *Phys. Rev. Letters*, **2,** 292.
Gold, A. V., 1958, *Phil. Trans.* A, **251,** 85.
Heine, V., 1957, *Phys. Rev.*, **107,** 431.
Langenberg, D. N., and Moore, T. W., 1959, *Phys. Rev. Letters*, **3,** 328.
Pippard, A. B., 1950, *Proc. roy. Soc.* A, **203,** 98; 1954, *Ibid.*, **224,** 273; 1957, *Phil. Trans.* A, **250,** 325.
Reuter, G. E. H., and Sondheimer, E. H., 1948, *Proc. roy. Soc.* A, **195,** 336.
Smith, G. E., 1959, *Phys. Rev.*, **115,** 1561.
Tegart, W. J. McG., 1956, *The Electrolytic and Chemical Polishing of Metals* (Pergamon Press).

The Anisotropy of Optical Absorption Induced in Sapphire by Neutron and Electron Irradiation†

By E. W. J. Mitchell, J. D. Rigden‡ and P. D. Townsend
Physics Department, The University, Reading

[Received April 27, 1960]

Abstract

Measurements have been made of the anisotropy of the optical absorption induced by neutron and electron irradiation. Some of the measurements were made with a six-plate fluorite polarizer, whose efficiency was found to be 65% at 2500 Å, and which could be used at shorter wavelengths than a Glan–Thomson prism.

In the spectrum induced by neutron irradiation, bands are resolved at −185°c at 1·9, 2·2, 2·8, 3·5, 4·1, 4·8 and 6·1 ev, the latter being measured both with a quartz spectrophotometer and a fluorite vacuum spectrograph. The 6·1 ev band was found to be isotropic but the 2·8, 3·5, 4·1 and 4·8 ev bands were anisotropic, having parallel to perpendicular ratios greater than 1. The anisotropy measurements reveal the presence of a band of opposite anisotropy at 5·3 ev. In order to account for the complete shape of the spectrum of the anisotropic ratio it is necessary to assume the presence of other bands having the same anisotropy as the 5·3 ev band. The 5·3 ev band did not appear in the electron irradiated specimens, although both the 4·8 ($\mu_{\parallel} < \mu_{\perp}$) and 6·1 ev (isotropic) bands were found. Possible models for these two centres are discussed.

The rate of decrease of absorption with heat treatment has been measured for temperatures up to 500°c and is compared with the rate of recovery of the neutron induced expansion studied by Martin (1959).

§ 1. Introduction

Both Levy (private communication, 1957) and Rigden (1958) have found that the radiation-induced absorption spectrum has more structure than originally reported by Levy and Dienes (1954). The spectrum comprises several overlapping bands which Levy has analysed into Gaussian bands having peaks at either: 6·02, 5·34, 4·84, 4·21, 3·54, 2·95 and 2·42 ev; or, 6·06, 5·82, 5·54, 4·84, 4·07, 3·44, 2·85 and 2·41 ev. It is clearly desirable to resolve these bands experimentally.

Sapphire has a rhombohedral unit cell and it is therefore likely that absorption bands associated with defects in the structure will be anisotropic. Different centres will, in general, be expected to show differing amounts of anisotropy. Thus, the study of the anisotropy of absorption, particularly at low temperatures, can be expected to give experimental information about some of the details of overlapping bands. We have

† Communicated by the Authors.
‡ Now at National Research Council, Ottawa.

therefore measured the anisotropy of absorption in polarized light from 6000–2000 Å.

It has already been shown by Gibbs *et al.* (1957) that a band can be produced at 4·8 ev by oxidation. In fact, this band is often found in unirradiated crystals, having presumably been formed in the oxidizing conditions prevailing at growth. However, before the 4·8 ev radiation induced band is ascribed, without doubt, to the same centre it is desirable that some property in addition to peak energy should be shown to be the same. We have used optical anisotropy as an additional property and the results are presented in the paper.

§ 2. Experimental

2.1. *Materials*

The crystals used were artificial sapphires grown by the Salford Electrical Instruments Ltd. Our specimens were in the form of discs 15 mm diameter and 1 or $\frac{1}{2}$ mm thick. The optic axis (c-axis)—that is, the body diagonal of the rhombohedral cell—is in the plane of the disc. Both this axis, and the perpendicular direction lying in the plane, were determined by back reflection Laue photographs. Spectrographic analysis indicated that Na (2 ppM), Ca, Cu, Mg and Si (1 ppM) were the principal impurities.

2.2. *Method of Measurement*

The optical transmission spectra were measured in two Hilger UVISPEK spectrophotometers, the useful ranges of which were about 10 000–2100 Å and 10 000–1950 Å respectively. Measurements in the vacuum ultraviolet were made using the apparatus described by Mitchell and Paige (1956). In the sapphire measurements, however, the plate calibration was determined using a set of sectored discs which were rotated in vacuum rather than by oscillating grids. The construction and operation of this device were described by Rigden (1958) and will be published separately.

Reflection losses were calculated using the refractive indices determined by Kebler (1954) and absorption coefficients were calculated from:

$$T = (1-R)^2 \exp(-\mu x) \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where T is the percentage transmission, R the reflection from a sapphire/air interface, μ (cm^{-1}) the absorption coefficient and x (cm) the specimen thickness. None of the crystals before irradiation had a transmission spectrum which could be accounted for by assuming that reflection was the only loss. The additional losses are believed to arise partly from surface scattering and in part from absorbing centres present in the crystals from growth. The former effect was eliminated as far as possible, by setting the crystal always in the same position in the mounting vice; this was particularly important in the bleaching experiments. In the unirradiated crystals the amount of absorption present was less than 3 cm^{-1} for photon energies up to 5 ev, but could be as high as 5 cm^{-1} at higher photon energies.

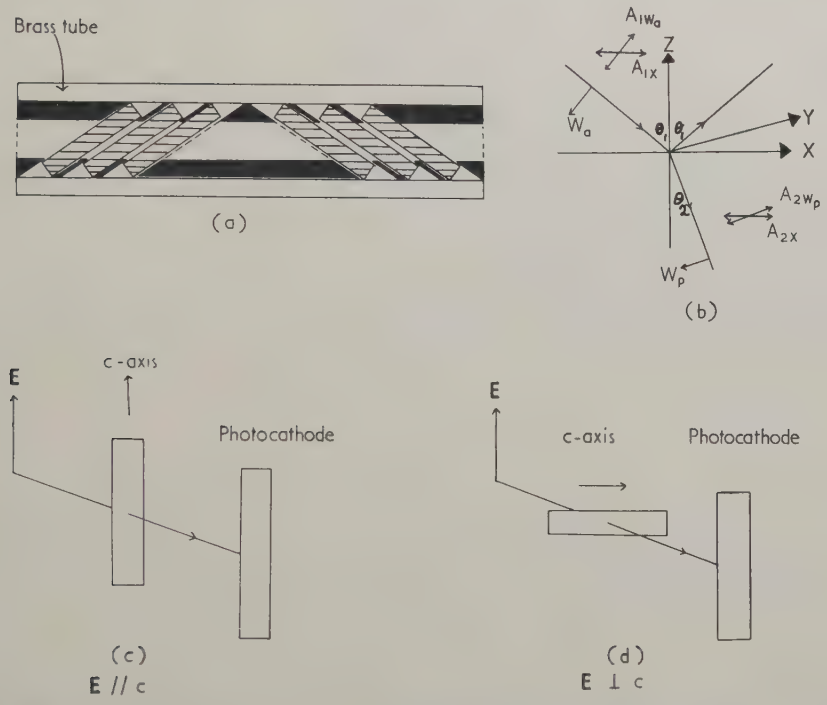
Much more absorption, up to 20 cm^{-1} at 6 ev, was present in one batch of crystals, which was not used in the irradiation experiments.

2.3. *Measurements using Polarized Light*

The quantity we wish to determine is the anisotropic ratio

$$(\mathscr{A} = \mu_{\mathbf{E}\parallel c}/\mu_{\mathbf{E}\perp c})$$

Fig. 1



(*a*) The fluorite polarizer—CaF_2 plates separated by brass spacers in a brass tube.

(*b*) Definitions of quantities used in calculating the efficiency of the polarizer: light in medium 1 incident at θ_1 and having amplitudes A_{1y}, A_{1x} and A_{1wa}; angle of refraction θ_2 and in medium 2 the amplitudes are A_{2y}, A_{2x} and A_{2wp}; the axes w_a and w_p are perpendicular to the directions of propagation in the respective media and lie in the plane of the beams.

(*c*) and (*d*) Relative positions of **E**, *c*-axis of crystal and photocathode surface, as used for determining $\mu_{\mathbf{E}\parallel c}$ and $\mu_{\mathbf{E}\perp c}$.

as discussed by Mitchell and Paige (1955), and Mitchell and Rigden (1957). Light from the monochromator is partially polarized and the response of the photocells is sensitive to the direction of **E**. Measurements of these effects were reported by Mitchell and Paige (1955).

In determining $\mathscr{A}$ it is necessary to eliminate the anisotropy of the sensitivity of the photocell. Thus we determine $\mu_\parallel$ and $\mu_\perp$ separately.

Having determined say $\mu_{\parallel}$ one could determine $\mu_{\perp}$ by rotating the polarizer through 90° and measuring $T_{\perp}$ in the usual way—by moving the specimen in and out of the beam. The disadvantage of this procedure is that if the initial ($\parallel$) direction is chosen to utilize the maximum intensity from the monochromator, the second ($\perp$) measurement is made with a smaller amount of energy, and therefore with less accuracy. Consequently we have determined ($\mu_{\perp}$) by rotating the crystal through 90°, and remeasuring the transmission curve in this position (fig. 1 (*c*) and (*d*)). Unless otherwise stated, the anisotropy before irradiation was found to be negligible.

2.4. *Polarizers*

Polaroid sheet was used as the polarizer in the visible region. For the ultra-violet a fluorite plate polarizer was made, some measurements being also carried out using a Glan-Thomson polarizer.

The plate polarizer contained six plates mounted in a brass tube with brass spacers as shown in fig. 1 (*a*). The plates were set at an angle (θ_1) of $(55{\cdot}5 \pm 0{\cdot}5)°$, which is the optimum polarizing angle for a refractive index of $1{\cdot}455 \pm 0{\cdot}014$ which for fluorite occurs at 2900 Å.

After crossing one interface the amplitudes of **E** in two directions y and w defined in fig. 1 (*b*), are given, for example in Ditchburn (1952), by:

$$A_{2w_p} = \frac{A_{1w_a} 2 \sin\theta_2 \cos\theta_1}{\sin(\theta_1+\theta_2)\cos(\theta_1-\theta_2)}$$

and

$$A_{2y} = \frac{A_{1y} 2 \sin\theta_2 \cos\theta_1}{\sin(\theta_1+\theta_2)} .$$

On emerging from the sixth plate the light is polarized by an amount represented by the polarization efficiency p, where

$$p = \frac{(A_{13w_a})^2 - (A_{13y})^2}{(A_{13w_a})^2 + (A_{13y})^2}$$

which can be written as:

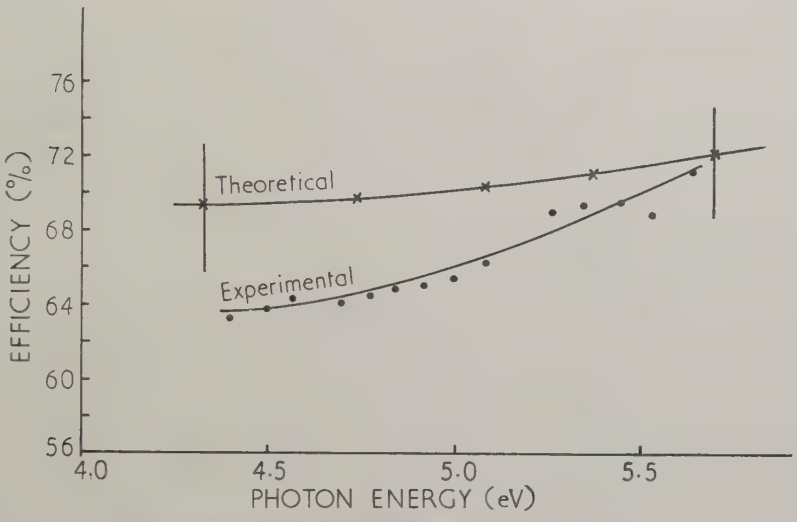
$$p = \frac{1-\cos^{24}(\theta_1-\theta_2)}{1+\cos^{24}(\theta_1+\theta_2)} . \qquad \ldots\ldots\ldots \quad (2)$$

The spectrum of the calculated efficiency is shown in fig. 2.

The spectral efficiency of the plate polarizer has been investigated using light polarized by the Glan–Thomson prism. The transmission of the plate polarizer ($T = I_T/I_0$) for both the parallel ($T_{\parallel}$) and perpendicular ($T_{\perp}$) orientations was measured in the UVISPEK. If the efficiency of the Glan–Thomson prism were 100% the quantity $p_e = (T_{\parallel} - T_{\perp})/(T_{\parallel} + T_{\perp})$ would represent the efficiency of the plate polarizer. The spectrum of p_e is shown in fig. 2, from which it will be seen that the measured efficiency is close to the calculated value, being 1 to 6% less throughout the range covered. This difference arises from the inaccuracy of setting the plates, the fact that the Glan–Thomson will not be 100% efficient and possibly some surface irregularities on the fluorite plates.

The use of a polarizer which is not 100% efficient means that the measured value of the anisotropic ratio will be less than the true value. For a true anisotropic ratio of 1·5 a 60% efficient polarizer will give a value of $\mathscr{A} = 1{\cdot}3$; and for a true value of 3·0 the measured value would be 2·0. In our measurements on sapphire a much larger uncertainty in the true value of $\mathscr{A}$ for a particular band arises because of the uncertainty about the relative amounts of the 'wings' of overlapping bands.

Fig. 2



Comparison of the measured and calculated efficiencies of the fluorite plate polarizer.

2.5. *Irradiations*

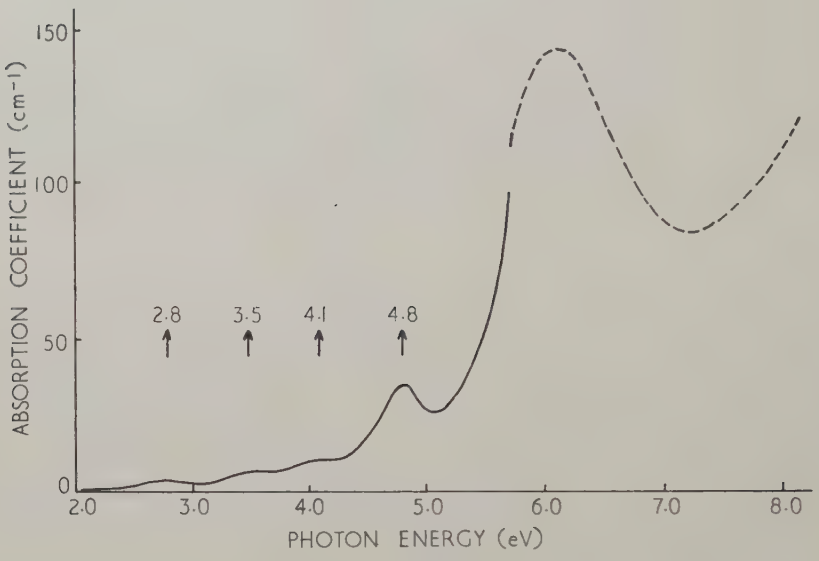
Neutron irradiations were carried out at about 30 °C in the Harwell pile BEPO. The specimens were irradiated with electrons from a 2 Mev Van de Graaff accelerator. For the electron irradiations the crystals were mounted in Wood's metal in a water cooled brass block.

§ 3. Results

3.1. *The Absorption Spectrum*

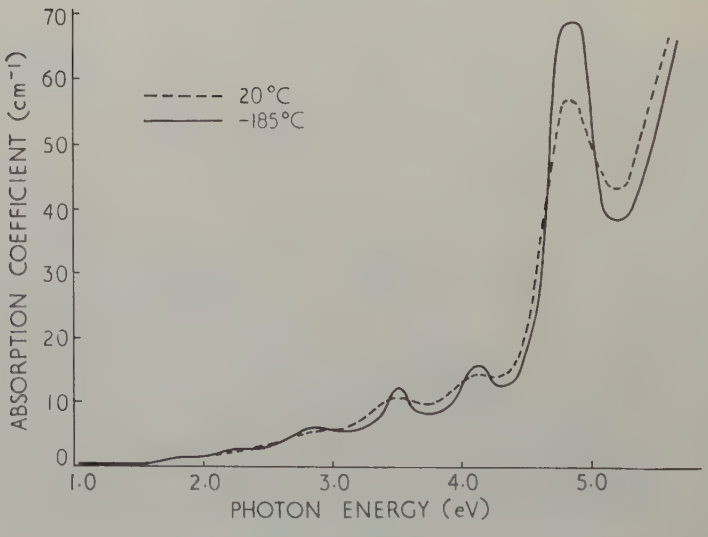
The spectrum measured at room temperature after reactor irradiation is shown in fig. 3. The measurements on this specimen for photon energies greater than 5·8 ev were obtained using the vacuum ultra-violet spectrograph and show the strong absorption band at 6·1 ev previously reported by Levy and Dienes (1954). For lower dose irradiations it is possible to see this peak in the spectrophotometer results. Other bands are found at 4·8, 4·1, 3·5 and 2·8 ev, the last three not being reported by Levy and Dienes (1954) but having been seen independently by Levy (1957). We shall refer later to the relative peak heights of the 6·1 and 4·8 ev bands; for the spectrum of fig. 3 the ratio of peak heights is 4·1 : 1.

Fig. 3



The absorption spectrum of neutron irradiated sapphire, including the vacuum ultra-violet region.

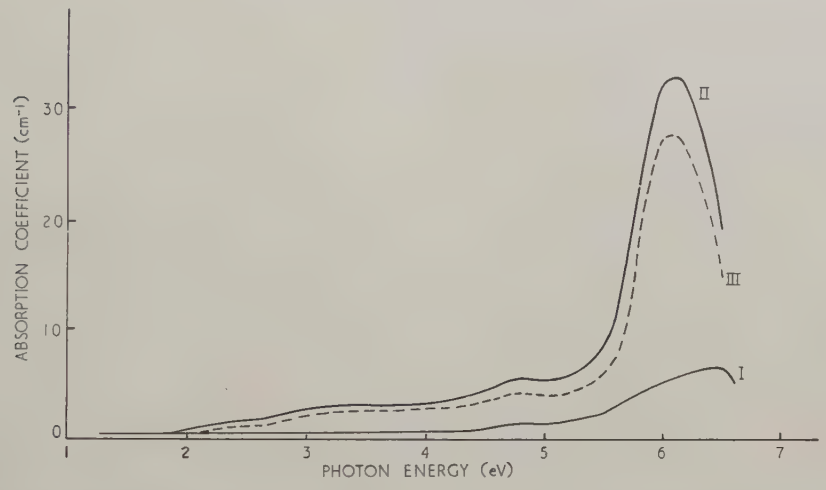
Fig. 4



The absorption spectrum of neutron irradiated sapphire at 20°c and −185°c.

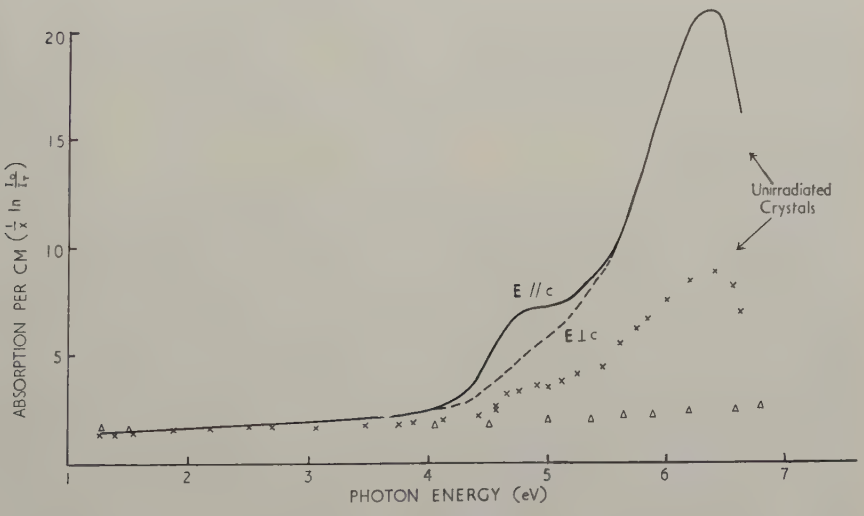
Low-temperature measurements have been made for photon energies up to 5·8 ev and as will be seen from fig. 4 the bands sharpen although the integrated absorption is within 3% of the absorption at room temperature. Additional bands are resolved at 1·9 and 2·2 ev. There was a trace of an

Fig. 5



The absorption spectrum before (I) and after (II) irradiation with 2 MeV electrons (10 500 μA min cm^{-2}); the induced spectrum is curve III.

Fig. 6



The absorption per unit length of two unirradiated crystals showing the 4·8 ev band. The points shown by △ △ △ are the calculated values of the reflection loss for a non-absorbing crystal.

absorption band at 1·9 ev in the x-ray induced absorption reported by Hunt and Schuler (1953), who found in addition weak (*ca.* 1 cm^{-1}) bands at 3·1 and 5·4 ev. Levy and Dienes (1954) also reported the presence of the 3·1 and 5·4 ev bands in γ-irradiated sapphire.

A typical spectrum induced by 2 Mev electron irradiation is shown in fig. 5. The strongest peaks of the spectrum after neutron irradiation are present (6·1 and 4·8 ev) and there is some indication of a band at 3·1 ev. It appears that the tail of the 6·1 band extends to the region of the 4·8 ev absorption. The ratio of the heights of the 6·1 and 4·8 ev peaks is 6·7. Because of the lower absorption this value has been determined from the induced spectrum and, as with the value quoted above for the neutron irradiation, no allowance has been made for overlap.

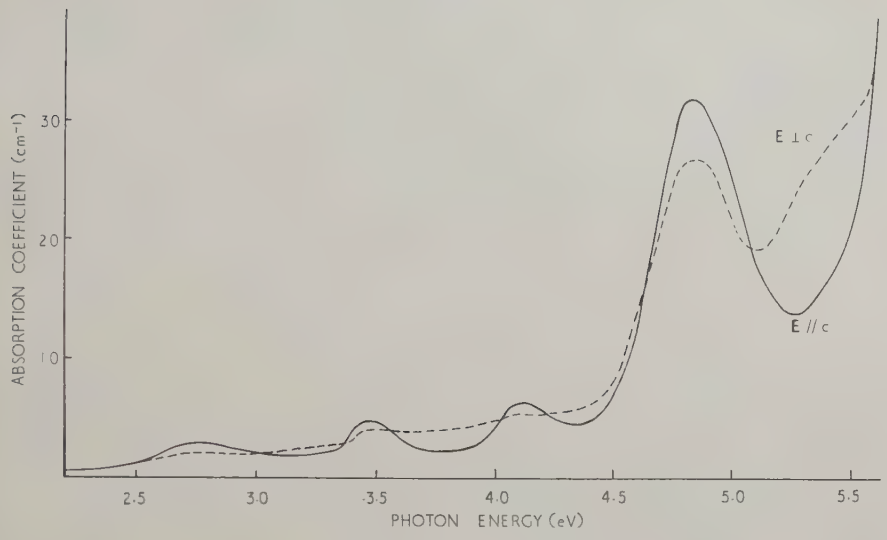
Gibbs *et al.* (1957) have reported in a brief communication that the 4·8 ev band could be produced in sapphire by heating in oxygen above 1300°c. They also found that the band could be removed by subsequent heating in hydrogen. We have found a weak band at 4·8 ev in crystals heated in air at 1400°c and the band was present before irradiation in one batch of crystals supplied by the manufacturer. The absorption spectrum of one of these crystals is shown in fig. 6.

3.2. *The Anisotropy of Absorption*

The anisotropy of the absorption induced by neutron irradiation is shown in fig. 7 and the spectrum of the anisotropic ratio ($\mathscr{A}$) in fig. 8, $\mathscr{A}$ for the 4·8 ev band being shown in more detail in fig. 9. The contribution from the 6·1 ev band appears to be isotropic and this has been confirmed at lower wavelengths using the fluorite plate polarizer. There are peaks showing positive anisotropy ($\mathscr{A} - 1 > 0$) at 4·8, 4·1, 3·5 and 2·8 ev. These peaks were all resolved in the straight-forward absorption spectrum. It will also be seen from fig. 8 that the anisotropy is negative for photon energies near 5·3, 4·4, 3·8 and 3·3 ev, none of which energies corresponds to resolved absorption peaks.

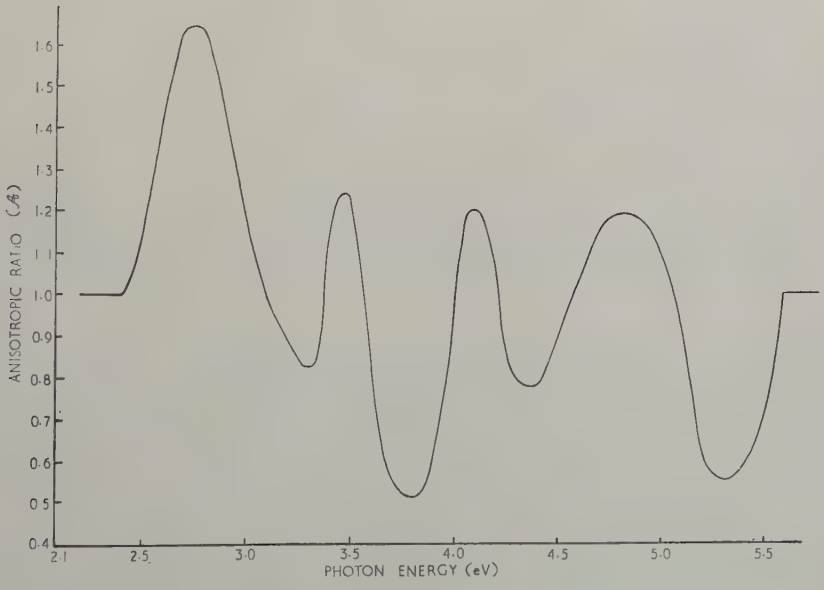
We consider first the negative anisotropy in the region of 5·3 ev. There is no doubt that the absorption at 5·3 ev is contributed to by the low-energy tail of the 6·1 ev band and the high-energy tail of the 4·8 ev band. However, the first of these is isotropic and the second has positive anisotropy so that to obtain a negative anisotropy we have to assume that there is a strong contribution at 5·3 from a band having negative anisotropy. The assumed band obviously cannot be sharp otherwise it would be resolved in the absorption spectrum. As was pointed out above Levy (1957) had to postulate a band either at 5·34 or 5·54 ev in order to account for the observed spectrum in terms of overlapping Guassian bands. Our anisotropic measurements are consistent with a band at 5·34 ev having a negative anisotropy such that $\mathscr{A} < 0{\cdot}55$. It is likely that this is the same band which is resolved in γ-irradiated specimens when both 4·8 and 6·1 ev bands are absent.

Fig. 7



Absorption spectrum induced by neutron irradiation: – – – – – – – – $\mathbf{E}_{\perp c}$; ———— $\mathbf{E}_{\parallel c}$. Measurements at −185°C.

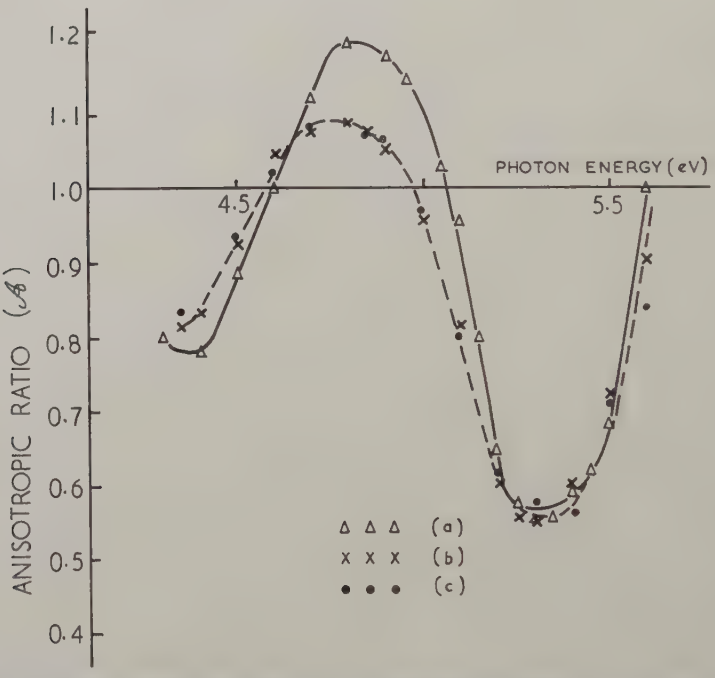
Fig. 8



Spectrum of the anisotropic ratio of a neutron irradiated crystal.

In order to account for the remaining part of the $\mathscr{A}$-spectrum it is not necessary to assume bands of negative anisotropy at all the minima (4·4, 3·8 and 3·3 ev). One simpler possibility is that there is a wide band having negative anisotropy at 3·8 ev. The small negative anisotropy at 4·4 ev could then be accounted for by overlap from the 3·8 and 5·34 ev bands (both negative) beyond the relatively narrower, and resolved, bands at 4·1 and 4·8 ev (both positive). Similarly, the small negative anisotropy at 3·3 ev could be accounted for by the stronger contribution of the negative 3·8 ev band beyond the narrower, and resolved, band at

Fig. 9



Spectrum of the anisotropic ratio in the region of 4·8 ev: (*a*) specimen R1 measured at −185°c using the Glan–Thomson; (*b*) R1 measured at 20°c using the fluorite plate polarizer; and (*c*) specimen R2A measured at 20°c with the plate polarizer.

3·5 ev (positive anisotropy). One alternative possiblity is that there is a wide unresolved 3·1 ev band present, the band which is found in γ-irradiated crystals. Such a band having negative anisotropy would account for the rapid decrease of $\mathscr{A}$ on the high-energy side of the 2·8 ev band (positive anisotropy), and would be sufficiently wide to account for the negative anisotropy at 3·8 ev, between the much narrower bands at 3·5 and 4·1 ev (positive anisotropy). For this alternative, however, one has to assume either an additional negative band at about 4·4 ev, or that the absorption at 4·4 ev is dominated by contributions from the 3·1 and 5·4 ev bands.

These results of measurements of the anisotropy of absorption may be summarized as follows:

Bands of positive anisotropy:

$$2{\cdot}8 \text{ ev } \mathscr{A} > 1{\cdot}65,$$
$$3{\cdot}47 \text{ ev } \mathscr{A} > 1{\cdot}24,$$
$$4{\cdot}10 \text{ ev } \mathscr{A} > 1{\cdot}20,$$
$$4{\cdot}80 \text{ ev } \mathscr{A} > 1{\cdot}20.$$

Bands of negative anisotropy:

$$5{\cdot}34 \text{ ev } \mathscr{A} < 0{\cdot}55.$$

At least one other $\mathscr{A} < 0{\cdot}50$,

either at 3·8 ev or 3·1 ev.

The positive bands are resolved as peaks in the absorption spectrum, the negative bands are not; in addition the isotropic 6·1 ev band is resolved in the absorption spectrum.

The anisotropy of the absorption spectrum induced by electron irradiation is shown in figs. 10 and 11. Because of the small absorption below 3 ev it was not possible to obtain experimentally reproducible results in that region. However, the fall of $\mathscr{A}$ below 3·4 ev is significant as are the values at higher photon energies. The most noticeable feature of fig. 11 is the absence of an effect which could be readily associated with a negative 5·34 ev band. The anisotropy spectrum could be accounted for by:

Bands of positive anisotropy:

3·4 ev $\mathscr{A} > 1{\cdot}6$,

4·8 ev (resolved in absorption spectrum) $\mathscr{A} > 1{\cdot}5$.

Bands of negative anisotropy:

Either 3·1 or 3·8 ev, or both, $\mathscr{A} < 1{\cdot}0$.

Neither of the negative bands is resolved in the absorption spectrum but the isotropic 6·1 ev band is resolved. We suggest that the overlap from the 3·8 ev band extends towards 5 ev and that the anisotropy of 1·5 at 5·3 ev is entirely due to overlap from the 4·8 ev band, there being also a contribution to the absorption from the isotropic 6·1 ev band. The 5·3 band if present in electron irradiated crystals, is very weak. Negative anisotropy at 5·3 ev was however obtained in such a crystal after subsequent neutron irradiation.

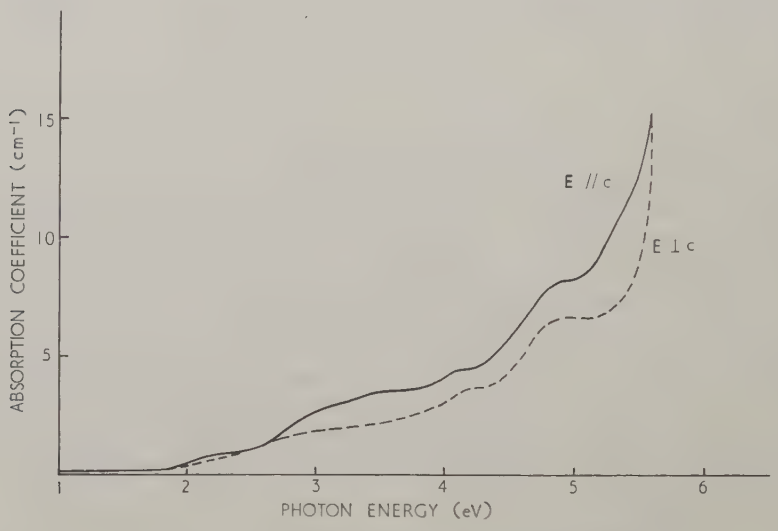
Finally the anisotropy has been measured of the absorption in one of the unirradiated crystals containing a band at 4·8 ev. The parallel and perperdicular curves are shown in fig. 6 from which it will be seen that the 4·8 ev band has positive anisotropy, as is found for the band induced at 4·8 ev by particle irradiation.

3.3. *The Effect of Heat Treatment on the Induced Absorption*

A neutron irradiated specimen was heated in an open furnace for 30 min at successively higher temperatures up to 500°c and the absorption spectrum

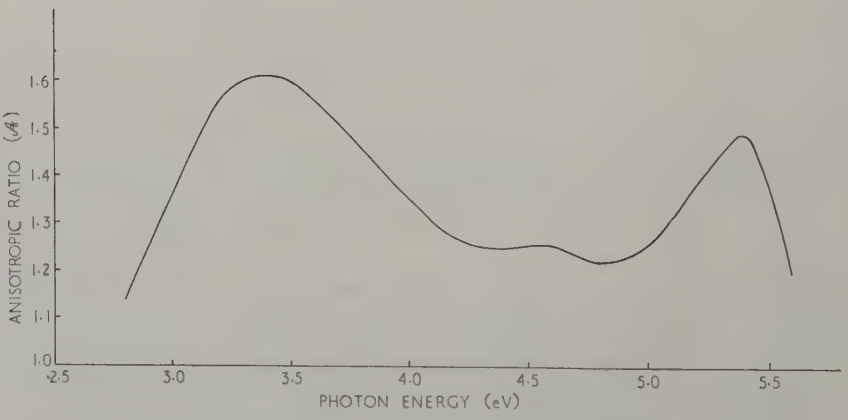
measured after each heat treatment. The isochronal recovery curves for the absorption peaks 4·1, 4·8 and 6·1 ev are shown in fig. 12, the results being normalized to be 100 before heat treatment. Martin's† (1959) results for the recovery of the length of reactor irradiated sapphire are also shown, having also been normalized to be 100 before heat treatment.

Fig. 10



The absorption spectrum after electron irradiation: ———— $\mathbf{E}_{\parallel c}$; and – – – – – $\mathbf{E}_{\perp c}$.

Fig. 11

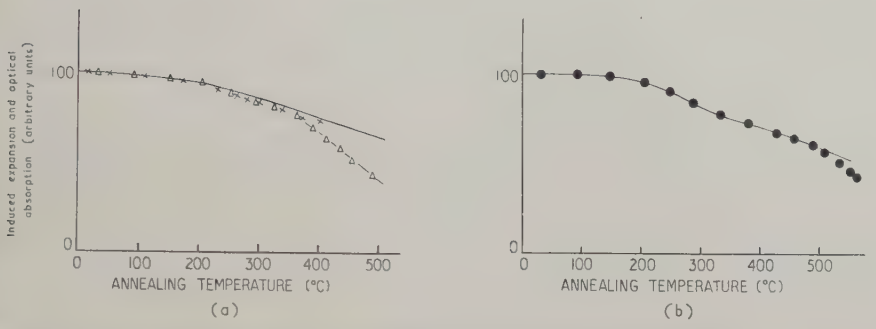


Spectrum of the anisotropic ratio of an electron irradiated crystal.

† The specimens used by Martin were supplied by the Salford Electrical Instruments, Ltd. and were reported to contain 3000 ppM of Si, 500 of Cr, 100 of Ni and 50 of Mn.

It will be seen that the rate of decrease of the 4·1 ev band is similar to the rate of recovery of the induced expansion parallel to the c-axis. On the other hand the 4·8 ev and 6·1 ev bands decrease at a rate similar to the rate of recovery of the *mean* induced expansion, although above 400°C the 6·1 ev band decreases more rapidly.

Fig. 12



The decrease of induced expansion (Martin) and optical absorption with heat treatment: (*a*) recovery of mean expansion shown as full line; (*b*) recovery of expansion ∥ c-axis shown as full line; optical absorption data shown by × × × (4·8 ev) and △ △ △ △ (6·1 ev) on (*a*); and ● ● ● (4·1 ev) on (*b*).

§ 4. Discussion of Results

The major additional feature which is resolved by measuring the anisotropy of the induced absorption is that at 5·3 ev, a band at this energy having previously been indicated by Levy's (1957) Gaussian analysis of the partially resolved spectrum. Because of the overlapping, however, it is not possible to determine exact values of $\mathscr{A}$ for individual centres, but only to fix lower, or higher limits.

The values of $\mathscr{A}$ to be expected for transitions involving wave functions localized in particular directions may be calculated from the relation given by Mitchell and Rigden (1957). For a uniaxial crystal having 3-fold symmetry about the axis:

$$\mathscr{A} = \frac{\sum_x 3n_x \cos^2\theta_x}{\sum_x n_x \cdot \frac{3}{2}\sin^2\theta_x}$$

where the set x of $3n_x$ dipole oscillators makes an angle θ_x with the c-axis. In the case where there is only one set of $3n$ at θ the anisotropic ratio $\mathscr{A} = 2\cot^2\theta$. Some values are given in the table for simple directions in the α–Al_2O_3 structure.

From the table it will be seen that using these directions the only way of obtaining small values of $\mathscr{A} < 1$ (negative anisotropy) would be from wave-functions directed along the closest 0–0 directions. In order to give non-zero values these directions would have to lie slightly out of the (001)

plane. Such a distortion could arise because of an adjacent Al vacancy. There are other possibilities for localized orbitals lying close to the (001) plane, e.g. the association of an Al vacancy and an interstitial impurity, or an oxygen vacancy with an impurity atom in a normal oxygen position. It is clear that the experimentally established negative anisotropy of the 5·3 ev band cannot be associated with Al–0 or Al–Al directions.

The anisotropy of the 2·8 ev band ($\mathscr{A} > 1{\cdot}65$) could be accounted for if the orbitals involved were directed either along Al–0 directions, or along some of the combinations of Al–Al directions, but not along 0–0. This band appears only after neutron irradiation and is presumably associated with a displacement, either in some stage of aggregation or single.

Anisotropic Ratios for Transitions Involving Particular Directions in Sapphire

Direction in crystal	Angle with c-axis	$\mathscr{A}$	Possible centre
Al–0	N.N. 3 at 46° 36′. Effect from N.N. and 3 N.N.N. at 59° 6′, assuming the 6 to be equivalent	1·79 1·16	Al vacancy; directed bonds from oxygens remaining in original directions (cf. Coulson and Kearsley 1957, diamond).
0–0	N.N. 3 at 90°. Effect from N.N. and 3 N.N.N. at 33° 48′	0 1·25	Al vacancy and a trapped hole; hole localized in N.N. 0—0 or shared along 6 edges of 0 octahedron
Al–Al	N.N. 1 at 0°. Effect from N.N. and 3 N.N.N. at 78° 25′.	∞ 2·33	0 vacancy and trapped electron shared between neighbouring Al–Al

The strongest band induced is that at 6·1 ev, which is isotropic. This centre could be an interstitial atom. If it were a substitutional centre the table shows that the transition could not be associated with the closest Al–0, Al–Al or 0–0 directions.

The positive anisotropy found for the 4·80 ev band, whether produced by irradiation or from growth is further evidence that it is the same centre involved in both cases. Gibbs *et al.* (1957) induced a band at 4·8 ev by heating in oxygen and we suggest that this is either due to interstitial oxygen, or the oxygen atoms surrounding an Al vacancy. However, the 4·8 ev band reported in this paper is not isotropic, so that we prefer the latter alternative. Finally, the 6·1 and 4·8 ev bands have always appeared together and with approximately the same relative strengths (5 ± 1). It is possible, therefore, that these two bands are associated with complementary centres. This would also be consistent with an interstitial model

for the 6·1 ev band, and leads to the specific suggestion that it results from an electron trapped at an interstitial Al ion. Arnold (1960) has also suggested, in a brief communication and for different reasons, that the 6·1 ev band is associated with Al displacements.

§ 5. Conclusions

The absorption spectrum induced by irradiation as measured with unpolarized light contains unresolved bands which can be detected using polarized light. The strongest damage band (6·1 ev) is isotropic having on its low-energy side a normally unresolved band (5·3 ev) which is only induced by neutron irradiation. Because the anisotropy of the 4·8 ev band has been found to be positive in both irradiated and unirradiated specimens it is concluded that this band is associated with the centre induced by oxidation as reported by Gibbs. On the basis of the present evidence we ascribe the 4·8 ev band to a transition on an oxygen ion surrounding an Al vacancy, and the 6·1 ev band to an electron trapped at an interstitial Al ion.

Independently of these models the decrease of optical absorption with heat treatment up to 400°c involves similar processes to the recovery of linear dimensions as reported by Martin (1959).

Acknowledgments

We wish to thank Professor R. W. Ditchburn for his interest and both Mr. J. W. Ryde, F.R.S., and Mr. L. A. Thomas of the G.E.C. Research Laboratories, Wembley, for arranging the supply of some of the crystals. Two of the authors (J. D. R. and P. D. T.) wish to acknowledge the receipt of D.S.I.R. maintenance grants.

References

Arnold, G. W., 1960, *Bull. Amer. phys. Soc.*, **5,** 167, IAl.
Coulson, C. A., and Kearsley, M. J., 1957, *Proc. roy. Soc.* A, **241,** 433.
Ditchburn, R. W., 1952, *Light* (London: Blackie),
Gibbs, P., Cutler, I. B., and Bates, J. L., 1957, *Bull. Amer. phys. Soc.*, **2,** 300, B12.
Kebler, R. W., 1954, *Proc. Inst. Soc. Amer.*, **9,** part 2. Paper No. 54–3–5.
Levy, P. W., and Dienes, G. J., 1954, *Report of Bristol Conference on "Defects in Crystalline Solids"* (London: Physical Society).
Martin, D. G., 1959, *Phys. Chem. Solids*, **10,** 64.
Mitchell, E. W. J., and Paige, E. G. S., 1955, *Phil. Mag.*, **46,** 1353; 1956, *Ibid.*, **1,** 1085.
Mitchell, E. W. J., and Rigden, J. D., 1957, *Phil. Mag.*, **2,** 941.
Rigden, J. D., 1958, Ph.D. Thesis, University of Reading.
Schuler, R. H., and Hunt, R. A., 1953, *Phys. Rev.*, **89,** 664.

Some Observations on the Stress-corrosion Cracking of α-brass and Similar Alloys†

By C. Edeleanu
Tube Investments Research Laboratories, Hinxton Hall, nr. Cambridge

and A. J. Forty
H. H. Wills Physics Laboratory, University of Bristol

[Received May 14, 1960]

Abstract

An attempt has been made to establish a mechanism for stress corrosion by direct observation on the propagation of cracks in α-brass tested under ammonia. The observations indicate that failure might be described by a two-stage process, of local embrittlement followed by a cleavage-like fracture. Such a model demands that the embrittled zone should have its mechanical properties modified in such a way as to allow a crack to form and reach some critical velocity, whilst the normal unattacked metal should be able to support this free-running crack for extensive failure to occur. These requirements are examined in the light of current ideas on the metal physics of fracture.

§ 1. Introduction

The failure of certain alloys under the simultaneous action of static stress and exposure to a corrosive environment is a well-known phenomenon (see, for example, the two recent symposia on the subject (Robertson 1956, Rhodin 1960)). There is as yet, however, no completely satisfactory explanation. The phenomenon presents an interesting relationship between mechanical failure and chemical attack, neither of which is sufficient alone to produce serious damage. Some explanations (for example, that put forward recently by Hoar and West 1958) emphasize the importance of the chemical attack, considering that the plastic yielding of the crystal under the applied stress enhances the rate of dissolution of metal at the tip of the crack. Others regard the failure as a purely mechanical breakdown in a material which has become embrittled by exposure to the corrosive reagent.

The present studies of crack propagation in stress corrosion were carried out in an attempt to distinguish between these processes by direct observation. We suggest below that the mechanical factor appears to be at least as important as the chemical one, although complete embrittlement is not essential ; local embrittlement of a zone at the free surface of a crystal or around the tip of a static crack is all that is

† Communicated by the Authors.

necessary to initiate fracture, and this can propagate through the non-embrittled metal provided a high crack velocity is maintained.

Our observations on the progress of cracks in α-brass lead us to propose a model of transcrystalline stress-corrosion cracking based on the process of embrittlement followed by brittle crack propagation. It is noteworthy that those systems which exhibit transgranular failure have a number of common features, all of which are required by this model. For example, alloys like copper/zinc and copper/gold probably contain varying degrees of short-range lattice order. This is significant since short-range order should have a marked effect on the mechanical properties of the alloy as far as crack propagation is concerned. Again, the corrosive environments which lead to cracking produce only a selective attack of the alloy (for example, ferric chloride oxidizes the copper in copper/gold, ammonia selectively removes zinc from brass). This is important since it can lead to a form of embrittlement which, according to our ideas, forms a possible mechanism for the initiation and further propagation of a brittle crack.

The observations on which this paper is based deal mainly with the *propagation* of cracks in stress corrosion. This arises from two considerations. Firstly, with present techniques it is not possible to observe the initiation of a crack directly ; the crack must have propagated some distance before it becomes visible. Secondly, we consider that the propagation of the crack is equally important in determining the failure of a specimen and this, of course, can be observed directly. Our experiments and the ideas developed from them do in fact justify this approach. It will be shown in a later section how these ideas tend to indicate the kind of relationship to be expected between susceptibility to stress-corrosion and alloy structure, composition, the degree of preliminary cold work and the nature of the corrosive environment.

§ 2. Experimental Technique

The system adopted for the present study is α-brass in ammonia. This is a useful system experimentally since α-brass is easily controllable in crystal form, composition and surface finish. We have used as our stock material a high purity brass supplied by I.C.I. (Metals Division), having an initial composition :

Cu	As	Sn	Ni	Fe	Mn	Bi
69·71	0·003	trace	0·01	0·01	0·05	trace

Some zinc is lost from this material during the preparation of single crystals but analyses have shown that this does not appreciably alter the composition, particularly in the central regions of the crystals where our observations of crack propagation are made.

Good single crystals can be grown fairly easily by using a closed graphite mould lowered slowly through a temperature gradient. A light mechanical polish of these crystals followed by electropolishing in a

phosphoric acid bath produces surfaces which are completely satisfactory for microscope observations.

Normally, when brass is exposed to ammonia vapour the surfaces tarnish very rapidly. This is due to the presence of oxygen and the tarnishing can be minimized by fully immersing the brass in a solution of ammonia. Consequently, in order to preserve the high optical quality of the surfaces, our observations have been made using an apparatus which allows continuous viewing at high magnification whilst the

Fig. 1

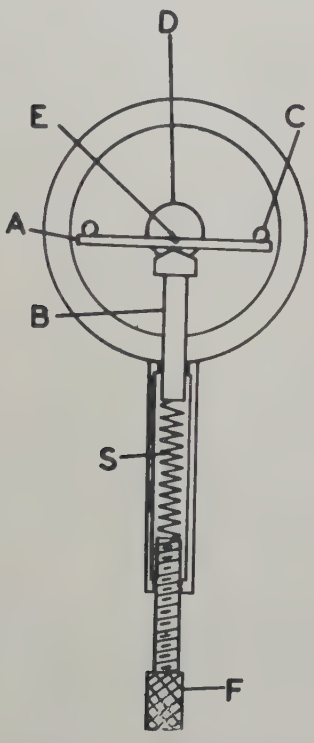


Diagram of straining apparatus. A. Brass specimen. B. Steel piston. C. Steel peg. D. Glass window (cover slip). E. Notch in specimen. F. Steel screw. S. Steel spring.

specimen is deformed in bending fully immersed in ammonia. A diagram of this apparatus is shown in fig. 1†. The load applied to bend the specimen is controlled directly by the compression of the spring under the movement of the screw.

The experimental procedure is as follows. The specimen is strained in bending with a residual stress applied through the compressed spring. Ammonia is added to the cell and the surface of the specimen is

† With the exception of figs. 1 and 7, all figures are shown as plates.

continuously viewed through the glass window with a high-power water-immersion microscope objective. Areas of the polished surface can be photographed before and after the crack appears and, in this way, it is possible to correlate the mode of propagation with the grain, orientation and pre-deformation of the specimen. This kind of observation is greatly assisted by notching the specimen before straining to localize the heavily cold-worked area and therefore the area of cracking. It must be emphasized that the notching is not an essential requirement for cracking to occur but is simply used for experimental convenience ; unnotched specimens usually fail by the formation of a large number of cracks, notched specimens by one main crack running from the notch.

The observations we have made in this way on the propagation of the cracks are obviously limited to surface traces. They have therefore been supported in a few cases by sectioning the specimens after cracking. This has confirmed that the main features of the surface traces, the discontinuities in these traces for example, are characteristic of cracking inside the crystal. That is, our observations, although confined to the surface, appear to provide a reliable indication of the mode of propagation of the complete crack.

§ 3. Experimental Observations

Failures by stress corrosion are usually classified as intergranular or transgranular cracking. With some materials, as for example with the particular α-brass used in the present experiments, there appears to be no fundamental difference between failures by these two paths. It is convenient however to describe the observations separately for single crystals (transgranular failure) and polycrystalline specimens (both intercrystalline and transgranular cracking).

Using the technique just described with single crystals it was noticed that the heavily cold-worked region appearing around the notch after the initial straining could not be prevented from tarnishing. This may be due to the increased chemical activity along slip lines in this area but entrapped air in the notch may also be partly responsible. The surface well removed from the notch is less heavily worked and less tarnished so that high-resolution microscopy is readily achieved here. The small amount of tarnishing which does occur appears to be selective and can in some cases be put to effective use for increasing the visibility of slip lines. For example, in fig. 2, which shows a surface crack initially formed by stress corrosion and then strained open by further stressing, the short slip steps around the tips of the crack are visible only as a result of the tarnishing. This micrograph provides at the same time a good illustration of the high degree of ductility remaining in the brass around the cracked region ; the fracture cannot propagate under stress alone. Moreover, localization of the slip at the tip of the crack indicates a pronounced stress-concentration effect which is characteristic of a mechanical failure,

and it is consequently difficult to believe that the crack has advanced in this state purely by chemical dissolution of the brass.

The appearance of the main crack at the notch invariably follows some period of incubation which under the present experimental conditions was about 1–2 hours. It is not possible to observe its formation or even its propagation in detail in this region. However, when the crack finally emerges into the less severely deformed part of the crystal a number of interesting features of the mode of propagation become apparent. These are best illustrated by sequences of micrographs like that of figs. 3–6.

Figures 3–6 show successive stages in the propagation of a free-running crack (i.e. no increased stressing during the test). The crystal in this case has an orientation such that the crack is running closely parallel to the trace of one slip system but cuts across the principal operative slip systems. The crack appears to run in short steps, each step starting and finishing on transverse slip lines. Some of these slip lines are apparently newly produced by the crack, others (and possibly all of them) are pre-existing slip bands which are increased in size by the crack. It appears then that rapid failure occurs in each step but this is terminated at pre-existing slip bands. The stresses around the tip of the crack are relieved now by further slip in the bands and this is accompanied by a corresponding amount of sidewise opening of the crack. This opening of the crack when it reaches a slip band is shown for example in figs. 3–6 and also diagrammatically in fig. 7 (A–C).

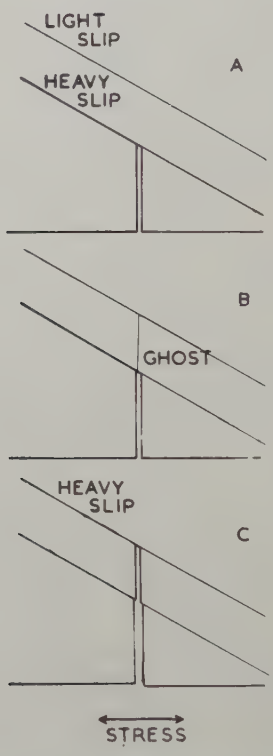
Although the crack opens mainly in the way described there appears to be also a slight ' yawning ' as well as a parallel displacement of its surfaces ; this is taken to indicate that the slip at its tip is being restricted or is non-uniform leading to a local rotation of the crystal structure.

The crack halts temporarily at each new slip band and there is a further period of incubation here, although this is not as long as that required for the appearance of the initial crack at the notch and may be as short as one minute. When the next stage in the fracture does occur it appears suddenly as a very fine crack, barely visible under the microscope. This we describe as a ' ghost ' crack. The ' ghost ', which terminates on a slip band, widens slowly into a visible crack by further slip. This succession of events strongly suggests that fracture is occurring, not by the propagation of a single crack, but by a discontinuous process involving the spreading, stopping and re-initiation of separate cracks. This discontinuous fracture process leads to an extremely irregularly shaped surface trace. In the particular case of figs. 3–6 we see that the trace is composed of short straight segments lying between neighbouring slip bands. Side-stepping of the trace occurs frequently ; this presumably means that the new ghost crack is re-initiated from some site further down the slip band or, alternatively, starts from the old crack but tends to spread initially along the slip direction. In either case, failure must eventually occur within the slip band otherwise the

full applied load could not be transmitted to the tip of the ghost to maintain propagation.

When the main crack direction is not as closely perpendicular to slip bands as in figs. 3–6, the crack pattern is less simple. In fig. 8, for example, the segments of the trace of the crack are curved but clearly some side-stepping occurs across most of the slip bands. The side-stepping is a characteristic feature of all the specimens tested and we infer that there must be some marked relationship between the propagation of the crack and the crystal structure, and in particular the occurrence of slip bands in the crystal.

Fig. 7



Diagrammatic representation showing stages during crack propagation.

The side-stepping is not confined to the surface traces. This has been demonstrated by taking transverse sections across the cracks. Figure 9, for example, is a micrograph showing in section one of the tested single crystals. We see that the discontinuous nature of the crack extends through the whole crystal ; the crack appears to be a rough fracture composed of a large number of cleavage-like elements. The orientation of the crack relative to the crystal structure in this particular case shows that the cleavage tends to follow the (111) plane.

Although it is not possible to discover the origin of the crack running from the notch, the origin of other minor surface cracking can sometimes be detected. This generally occurs where duplex slip has taken place. That is, where slip is held up by other slip bands, a sufficient stress can be concentrated here to start a crack under the ammonia. A large number of sites of this kind can be seen in fig. 10. Figure 11 shows how, alternatively, cracks can be formed inside heavy slip bands. Some caution must be applied however in attributing all forms of cracking along slip bands to a failure of the bulk material ; some cases have been found in which rupturing the film of tarnish on the surface as the slip steps emerge could be mistaken for cracking of the metal itself.

A number of experiments have been carried out on polycrystalline brass generally prepared by annealing the cold-rolled stock in air at 800°c. We have found that the cracking occurs predominantly near grain boundaries but that a great deal of transcrystalline cracking is also present. The careful study of intercrystalline cracking using the present technique is complicated by the fact that a slight amount of boundary etching is often present, making the detection of the ghost difficult. Observation is further complicated by the branching generally found and the difficulty of predicting which of a number of possible boundaries is likely to fail. However, such evidence as has been obtained indicates that the cracking occurs in exactly the same way as in the single crystal except that it follows a boundary instead of a cleavage plane. As in the single crystal case quite apart from the main cracking from the notch, surface cracking also is found, but in this case it is often associated with boundaries and is considerably more frequently observed.

In many cases this surface cracking develops from a series of micro-cracks formed in the neighbourhood of boundaries. That is, the cracks following the grain boundaries do not appear to be produced by a simple parting of the grains. The micro-cracks are formed at the boundaries but extend into the adjacent grains under the applied stress. This kind of failure is illustrated by figs. 12, 13 and 14. The appearance of the micro-cracks in fig. 12 strongly suggests that they might be formed by the mechanism proposed by Stroh (1957); that is, the grain boundaries sufficiently oppose slip within the grains for stresses to be concentrated here to initiate fracture. Both the correlation between the occurrence of slip lines and cracks and the shapes of the cracks agree with Stroh's ideas on the initiation of fracture in a ductile material. It should be noted however that these cracks appear only after immersion in ammonia. This implies that, since the concentration of stress at the head of slip band is not normally great enough to cause fracture under the kind of stresses applied during the test, the corrosive environment is playing an important part in modifying the mechanical properties of the crystal in the neighbourhood of grain boundaries to allow this early failure.

The micro-cracks spread under the applied stress and eventually link up to complete the intercrystalline breakdown. The inclination

of the boundary to the principal stress direction seems important in determining the ease with which the eventual link up proceeds, as can be seen in figs. 13 and 14. These show two examples where boundaries lying in different orientations have reached different stages in this process. Once the micro-cracks have linked up completely to form a large intercrystalline failure it is very easy to overlook this detail in reconstructing the mechanism by which the cracking started. It may well be that this explains why this effect has not been previously reported.

§ 4. Discussion

The most important impression gained from the microscope observations is that failure of the specimen occurs by a typical fracture process ; that is the cracking is largely mechanical in character. Yet this kind of failure occurs only in the presence of ammonia. Without the ammonia or with a sudden increase of stress the crystal deforms plastically around the tip of the crack. This has been illustrated by fig. 2 and also by a cine record of the cracking process. The phenomenon of stress-corrosion cracking then might well be the development of a brittle-like fracture triggered in some way be a chemical reaction between the brass and the ammonia.

The complete failure of the specimen takes a finite time, greater than that required for a purely brittle fracture. It is apparent however that the individual steps in the cracking process occur very rapidly and might well be brittle fracture whilst the process of re-initiation of the cracks at slip bands is a slow process leading to a low overall failure rate. In other words the rate-controlling process is the re-initiation of the crack. As we shall see later this re-initiation might well be the most important aspect of the whole phenomenon of stress-corrosion cracking. Our interpretation of the phenomenon then is that a crack runs freely through the brass in a brittle fashion until it meets a slip band. Slip in this band stops the crack. Further cracking requires re-initiation of the fracture and this, being associated with a chemical process, takes time and therefore controls the rate of the whole process.

Such a model requires in the first instance that undeformed α-brass can support a free-running crack. A truly brittle fracture is possible only if the plastic yielding of the metal under the stress field imposed by the crack can be restricted. Previous work on this aspect of brittle fracture in ductile crystals (Gilman 1957, 1958, Forty 1957) has indicated that this is possible only if the crack travels faster than some critical speed which is associated with the rate at which dislocations can be generated and moved outwards from the tip of the crack. We have no measure of the speed at which the brittle bursts of fracture propagate in stress corrosion but it is possible to put some upper limit on it since it is unlikely that any brittle fracture will spread at a speed greater than one-third of the speed of sound in the crystal. It follows that if yielding is to be prevented dislocations in α-brass must move slower

than this. This is in fact quite possible since it is well known that α-brass reacts sluggishly to applied stress, this being attributed to the short range ordering of the structure characteristic of all alloys of this kind (Cottrell 1953, Broom and Biggs 1954). Copper/gold, austenitic stainless steel and other materials susceptible to stress-corrosion cracking all have a characteristically low yield rate which can be attributed to the interaction of dislocations with phase domains, or impurity atmospheres.

The interpretation of our observations on the propagation of cracks in terms of short bursts of brittle fracture must clearly arouse some interest. It implies that brittle fracture is possible in a face-centred cubic metal and, moreover, that it is a cleavage-like fracture which appears to follow the {111} planes. The foregoing discussion has shown that this can be justified only if the crack is propagated at high speed. We suggest that the important role of the chemical environment in stress corrosion of these alloys is to modify the mechanical properties of a surface layer or a zone around the tip of an exhausted crack in such a way not only to allow a new crack to form but also to reach the critical speed required for further propagation in the normal metal.

Our model also demands that deformed α-brass should be softer than the undeformed material. As we have seen, the crack is free-running until it meets a slip band, where it stops by a process of further plastic flow. This too can be explained on the basis of the existence of short-range order in our specimens ; for there is good reason to believe that the localized destruction of this order in the slip bands does in fact lead to a form of softening (as demonstrated, for example, by the yield point phenomenon in copper/gold). Softening can occur only initially and the slip bands must eventually work-harden in the normal way. This suggests that a small degree of cold work might be effective in stopping the cracks and hence possibly decreasing the susceptibility to stress corrosion, whereas heavy cold work might support cracking.

Our model proposes that, when a crack terminates at a slip band, fracture continues only by re-initiation and propagation into the next undeformed zone of the crystal. We have examined a number of mechanisms for the process of re-initiation, which may differ from alloy to alloy. One such mechanism is based on the idea that, although the shear stresses at the tip of the incident crack are relieved by slipping in a pre-existing slip band, this might not completely remove stress concentrations since the slip can be held up further down the band. If the obstacle here is strong enough a new fracture can occur. It is difficult however to understand how such a mechanism is influenced by the chemical environment, unless this aids the slip process at the tip of the crack and therefore indirectly assists the concentration of stresses inside the crystal.

An alternative mechanism, and one which is apparently more satisfactory since it utilizes the chemical environment more directly, is that of embrittlement. If the corrosive reagent produces embrittlement

of the metal around the tip of the crack, then clearly a purely brittle fracture can be initiated here. Embrittlement can occur also over the surface of the specimen in early stages of the test and thus account for the initiation of cracks in the first place as well as for their subsequent re-initiation as the failure spreads. Complete embrittlement of the specimen is not essential and indeed is not reached in practice according to the observation that marked plastic deformation can be produced by an increase in stress or by removal of the ammonia during test. It is necessary only for an embrittled zone to be formed at the free surface or around the tip of an existing crack. The zone must be large enough however to support a crack which is stable at the stress level being applied and must, moreover, allow the crack to reach a critical velocity so that the brittle fracture can be maintained when it emerges from the embrittled zone and penetrates the normal metal. The critical size, according to current ideas of the cracking of brittle solids, is likely to be inversely dependent on the square of the stress applied to open the crack. On this account a very marked effect of applied stress on time to failure can be expected, which agrees with experiment (Robertson 1956).

The growth of the embrittled zone to critical size will of course account for the incubation periods found for the initiation of the original crack at the surface and for the re-initiation after its arrest at slip bands. Some evidence for the existence and growth of the zone may be derived from figs. 15 and 16. These show the results of increasing the bending stress during a test, immediately after the halting of a crack in one case (fig. 15) and after a waiting period in the other (fig. 16). We see that the increased stress is accommodated by plastic yielding around the tip of the newly halted crack in the first case but by a further burst of cracking in the other.

The localization of surface cracking around grain boundaries in polycrystalline specimens can also be accounted for by our model. It is likely that chemical reaction between the brass and ammonia would proceed more readily in the neighbourhood of a boundary and in this way a zone of embrittled material can develop here. This can restrict the slip within the grains and also resist the tendency for plastic deformation to be transferred from one grain to another so that micro-cracks may be more readily formed. There are materials, however, which fail in a truly intercrystalline fashion. In such cases a precipitation reaction is normally involved and this will affect the local mechanical properties and may also localize the corrosion in the region of grain boundaries. A good example of this class is the Al–7% Mg type of alloy.

The way in which embrittlement is produced by the chemical environment is not yet clear. It is known of course that brass can be embrittled by the process of dezincification during a vacuum anneal at high temperature, and an intergranular failure very similar to that of stress corrosion can be produced in a stressed specimen by this treatment. It is possible that a similar form of dezincification occurs when brass is

immersed in ammonia. However, the electrochemistry of the brass/ammonia system is incomplete and only simple tentative results can be quoted here.

A few potential measurements have shown that zinc, as expected, is considerably more anodic than copper in ammonia and in ammonia containing copper and zinc in the 70/30 ratio. It is suggested therefore that there will be a tendency for the selective removal of zinc from the alloy especially in the strictly limited volume of solution present in a crack. This leaching process may not be merely superficial and zinc could be removed from internal lattice sites by a dislocation piping mechanism. This would leave a partly porous metal since large-scale rearrangement by diffusion is unlikely at the temperature of the test. Such material would be well suited for the initiation of a brittle fracture. This idea has been developed further in a recent publication by Forty (1960).

Evidence for embrittlement in the copper gold alloys has already been provided by the work of Robertson and Bakish (1956). This shows that the selective removal of copper by ferric chloride solution (which ultimately leads to stress corrosion cracking) leaves behind a gold sponge. These authors make the further interesting observation that immersion of a slab of material in ferric chloride for several weeks results in complete penetration by tunnelling, possibly along dislocation lines.

An alternative explanation for the formation of the embrittled zone is that both components of the alloy (Cu and Zn in our case) are removed by the ammonia but that one of them (Cu possibly) is re-plated to form an electro-deposited sponge. One difficulty here however is that the sponge in this case might not be sufficiently coherent with the unattacked brass to allow transfer of the fracture into the normal ductile material. In other words a piped embrittled zone has more attractive mechanical properties for our requirements than a sponge.

The picture we have of the phenomenon of stress corrosion cracking can now be summarized. Exposure to a corrosive environment produces a surface embrittlement. Eventually, when the embrittled zone has reached a sufficient depth this layer is cracked by the applied stress and, whenever the crack reaches a high enough speed, cracking can continue in the brass until a soft region such as a slip band is reached. Here the crack is stopped by slip occurring at its tip. Further chemical embrittlement now occurs at the tip of crack until the embrittled zone is large enough to allow a new crack to form and reach the critical velocity for fracture. The overall rate of propagation of the crack and therefore the susceptibility of the material to stress-corrosion cracking is effectively controlled by the time between halting and the initiation of the new crack, and this in turn depends on the rate of chemical embrittlement around the tip and the magnitude of the applied stress. Thus we expect in general a relationship between susceptibility to failure and applied stress. The mechanism of embrittlement is not known but it seems

reasonable that the embrittled material will have mechanical properties strongly dependent on composition and state of deformation of the alloy. There should, therefore, be a relationship between susceptibility and composition. There seems to be some evidence for this in practice (Graf 1956). Further, the rate of cracking and therefore susceptibility should depend on the frequency with which soft slip bands are encountered. That is, susceptibility should depend on the degree of cold work in the specimen before testing.

One of the most important questions raised by this work is whether an undeformed crystal can support a free-running crack. This is only likely if the rate at which plastic flow can occur is small compared with the crack velocity. As we have already indicated, although this is impossible for pure metals, it might well be possible for certain types of alloy, and in particular for those alloys known to be susceptible to stress corrosion.

Acknowledgments

We wish to thank the Director of the Hinxton Hall Laboratory, Dr. T. P. Hughes, and the Chairman of Tube Investments Limited for permission to publish this work.

It is a pleasure also to thank Mr. T. Law for his skilful assistance during the course of the experiments.

References

Broom, T., and Biggs, W. R., 1954, *Phil. Mag.*, **45,** 246.

Cottrell, A. H., 1953, *Relation of Properties to Microstructure* (Cleveland : American Society for Metals), p. 131.

Forty, A. J., 1957, *Proc. roy. Soc.* A, **242,** 392 ; 1960, *Physical Metallurgy of Stress Corrosion Fracture*, ed. T. N. Rhodin (New York : Interscience Publishers), p. 99.

Gilman, J. J., 1957, *Trans. Amer. Inst. min. (metall.) Engrs*, **209,** 449 ; 1958, *J. appl. Phys.*, **29,** 601.

Graf, L., 1956, *Stress Corrosion Cracking and Embrittlement*, ed. W. D. Robertson (New York : J. Wiley & Sons).

Hoar, T. P., and West, J. M., 1958, *Nature, Lond.*, **181,** 835.

Rhodin, T. N., 1960, *Physical Metallurgy of Stress Corrosion Fracture*, ed. T. N. Rhodin (New York : Interscience Publishers).

Robertson, W. D., 1956, *Stress Corrosion Cracking and Embrittlement*, ed. W. D. Robertson (New York : J. Wiley & Sons).

Robertson, W. D., and Bakish, R., 1956, *Acta Met.*, **3,** 292.

Stroh, A. N., 1957, *Advanc. Phys.*, **6,** 418.

Thermal Aspects of the Growth of Thin Films by Vacuum Sublimation†

By G. Gafner
National Physical Research Laboratory,
P.O. Box 395, Pretoria, South Africa

[Received May 9, 1960]

§ 1. Introduction

It has been stated (Pashley 1956) that a knowledge of the growth mode of a layer is essential before any comprehensive theory of epitaxy can be formulated. Considerable experimental information regarding the initial stages of the overgrowth process has recently been published but interpretation of these results is hampered by the absence of quantitative information regarding the thermal situation prevailing during growth. It is to be expected that the growth temperature would markedly influence such important properties as crystallite size and perfection and the tensile stresses which might result if any differential thermal contraction of the substrate and overgrowth took place.

The following contradictory opinions are found in the literature regarding the temperature of an overgrowing layer: Hoffmann *et al.* (1954) have stated that the first and second nearest neighbours of an incident atom will undergo an instantaneous temperature rise of about 700°C when the incident atom shares its sublimation energy with them. As the relaxation time of such a thermal pulse is about 10^{-12} sec they conclude that it is incapable of producing any bulk recrystallization. This argument is undoubtedly valid for the incidence of an isolated atom but the conclusion is controversial when the film is heated by the continuous deposition of material. In contradistinction to this, Wilman (1955) has concluded, on the basis of experimental results, that the surface temperatures of overgrowths formed on metallic and other substrates at rates of 15 to 300 Å/sec may be hundreds of °C higher than the substrate temperatures. The need for a quantitative estimate of this temperature is thus apparent.

§ 2. Factors which Complicate the Mathematics

An exact mathematical representation of the thermal situation during the growth of a continuous film by vacuum sublimation is analytically

† Communicated by the Author.

intractable due to the presence of the following complications:

1. A free boundary with continuous heat generation on it.
2. Radiative heat losses from both the free surface and the substrate underface which depend on the fourth powers of the respective, varying temperatures.
3. The composite nature of the system as two different sets of thermal constants are involved; one for the substrate and another for the overgrowth.

It follows that any analytical treatment of the problem must involve many approximations. It will be shown that it is nevertheless possible to approximate to the various situations in such a way that reliable conclusions can be drawn.

§ 3. Range of the Substrate and Deposition Parameters

The range of final film thicknesses and deposition rates of interest have upper limits of 10 000 Å and 1000 Å/sec respectively. The substrate thicknesses vary from about 1 cm, in lens coating work to $< 0{\cdot}01$ cm when deposition is onto thin mica cleavages. When the substrate is thick it is to be expected that the film temperature will be predominantly dependent on the conduction losses across the interface. When the substrate is thin, however, the heat generated will be rapidly distributed through the composite system and its resultant temperature will depend on the balance between radiative losses and the heat generation rate. As the total amount of heat liberated (about 2 cal), when 1 cm^2 of the thickest film of interest (10 000 Å) is formed, can uniformly heat a substantial substrate only a few °C, it is reasonable to assume a semi-infinite substrate in this instance.

§ 4. Temperature of a Continuous Layer Growing on a Thick Substrate

An upper limit to the temperature in a film growing on a semi-infinite substrate will follow from a calculation of the surface temperature of the substrate when the heat associated with the deposition rate is generated directly on it. To ensure that this calculation gives a true upper limit, the lower of the substrate and overgrowth thermal diffusivities will be accredited to the substrate and no heat losses from the system will be allowed.

The heat-flow theory of this system is given by Carslaw and Jaeger (1948, p. 56) and the relevant special case of their solution is:

$$v = \frac{2Q}{K}\left(\frac{\alpha t}{\pi}\right)^{1/2} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

This equation relates the temperature increase v at a time t relative to that at time $t = 0$, on the surface of a semi-infinite solid on which heat

is generated at a constant rate Q cal/cm^{2}sec subsequent to time $t=0$. K and α represent the thermal conductivity and diffusivity of the solid respectively. For solids, using the c.g.s. °c system with the unit of heat the calorie, as will be done in all the ensuing work unless otherwise stated, $\rho c \approx 0{\cdot}6$ and thus $K \approx 0{\cdot}6\alpha$. This being so (1) becomes:

$$v = \frac{2Qt^{1/2}}{0{\cdot}6(\alpha\pi)^{1/2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The highest heats of vaporization lead to a heat generation rate of about 0·2 cal/cm^{2}sec when growth takes place at a rate of 1000 Å/sec. The lowest thermal diffusivity values for solids are around 0·005 cm^2/sec, so that the highest temperature rise to be expected for any film growing at a rate of 1000 Å/sec or less, on a thick substrate, to a thickness of 10 000 Å is, from (2):

$$v_{\max} = \frac{2 \times 0{\cdot}2 \times 10^{1/2}}{0{\cdot}6 \times 0{\cdot}0157^{1/2}} = 17\,^{\circ}\text{c}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The general conclusion is thus that there will be no marked temperature rise at the surface of a continuous, growing, film due to the build-up of heat at the plane of generation. This conclusion is valid for all materials commonly used in vacuum sublimation work.

§ 5. Temperature of a Continuous Layer Growing on a Thin Substrate

An implication of the previous conclusion is that large surface temperatures can arise only when the thermal capacity of the substrate is small. This in turn implies that there will be a small temperature difference across the substrate. When this situation prevails, an estimate of the temperature can made as follows:

Consider unit surface area of a substrate of the infinite slab type, which has thermal capacity A and is sufficiently thin to allow the temperature drop across it to be ignored when heat is continuously generated on one of its faces at a rate Q. This heat generation is associated with the advent of new material at such a rate that the thermal capacity of the system increases at a constant rate B. The thermal capacity of the system at time t is then given by $(A+Bt)$.

Heat losses are considered to be solely radiative and can thus be represented by a term of the form $C(v^4-v_0^4)$, where v is the varying temperature of the system and v_0 is the environmental and initial substrate temperature in °k. The net heat gain rate when the temperature is v is equal to $Q-C(v^4-v_0^4)$. The rate of change of temperature with time at any instant is thus:

$$\frac{dv}{dt} = \frac{Q-C(v^4-v_0^4)}{A+Bt}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The solution to this differential equation is readily found, after separation of the variables, and is:

$$\arctan(zv) - \arctan(zv_0) - \frac{1}{2}\left\{\log_e\left(\frac{zv-1}{zv+1}\right)\left(\frac{zv_0+1}{zv_0-1}\right)\right\} = 2z\left(\frac{Q+Cv_0^4}{B}\right)\log_e\left(1+\frac{Bt}{A}\right) \qquad (5)$$

where

C = Stefan–Boltzmann constant × number of faces × emissivity
$= 1{\cdot}37 \times 10^{-12} \times 2 \times$ emissivity;
A = Initial thermal capacity of the system/cm² of surface area
$= l_s\rho_s c_s$ (in which subscript s refers to the original substrate);
B = Thermal capacity of the material deposited/cm² sec
$= l_0/\text{sec}\,\rho_0 c_0$ (in which subscript 0 refers to the overgrowth);
Q = Heat released/sec cm² when the overgrowth thickness increases by l_0/sec;
v = Temperature of the system (°K);
v_0 = Environmental and initial temperature (°K);
$z = (Q/C + v_0^4)^{-1/4}$.

As eqn. (5) does not contain the temperature explicitly it is necessary to calculate the times at which given temperatures will be attained in order to obtain the temperature–time dependency. This has been done for a range of deposition rates, a substrate thickness of 1 mm and various substrate emissivities, on the assumption that a deposition rate of 1000 Å/sec corresponds to a heat generation rate of 0·2 cal/cm² sec and that the environmental temperature is 300°K. The results of these calculations are given in table 1.

Table 1. Times taken for the surface temperature of a thin slab to attain the stipulated values at various deposition rates and substrate emissivities

Heat generation rate in cal/sec cm²	Emissivity	Deposition in Å/sec	Time in seconds of attainment of			
			400°	500°	600°	700°K
0·2	0·02	1000	30	61	92	123
0·02	0·02	100	307	638	1031	1459
0·002	0·02	10	Max. temp. attained 303°K			
0·2	0·2	1000	31	65	104	285
0·02	0·2	100	590	Max. possible temp. 460°K		
0·002	0·2	10	Trivial temp. rises			

The times at which the temperatures listed in table 1 will be attained for different substrate thicknesses can be deduced from the table by multiplying the times in the table by the thickness of the substrate in mm. It follows that a temperature rise of 100°c or more, for example, will be attained during the deposition of a film of thickness 10 000 Å only if the substrate has a thickness of 0·03 cm or less, when the deposition rate is 1000 Å/sec and the substrate emissivity is 0·02. A maximum substrate thickness of 0·003 cm will allow the same temperature increment to be attained when the deposition rate is 100 Å/sec and the emissivity is as above. Increasing the emissivity of the system to 0·2 reduces the last mentioned maximum substrate thickness to 0·002 cm. It can thus be concluded that temperature increases of hundreds of °c will not arise unless deposition is onto very thin substrates.

§ 6. Validity of the Thick and Thin Substrate Approximations

The temperature of a continuous overgrowing layer has been calculated in approximation using two extreme models. The range of validity of these models can be estimated from a knowledge of the temperatures of, and the temperature difference between, the two substrate surfaces when heat is generated directly onto one of these surfaces and no radiative heat losses occur from either. Macey (1940) has shown that the temperature v in an infinite slab, measured with respect to the initial temperature, when heat is generated at a constant rate Q/unit area sec at $x=l$ and there is no heat loss from $x=0$, is given by

$$v=\frac{2Q(\alpha t)^{1/2}}{K}\sum_{n=0}^{\infty} i\operatorname{erf} c\frac{(2n+1)l-x}{2(\alpha t)^{1/2}}+i\operatorname{erf} c\frac{(2n+1)l+x}{2(\alpha t)^{1/2}}.$$

The temperatures at $x=0$ and $x=l$, designated v_0 and v_l respectively, have been calculated for various substrate thicknesses, and for times up to 10 sec when the heat generation rate is equal to the extreme value of 0·2 cal/cm² sec, taking the thermal diffusivity as 0·005 and the thermal conductivity as 0·003 both of which are extremely low values for solids. The results of these calculations are given in table 2.

These results show that the temperature differences across all substrates thinner than 1 mm are trivial. The theory which in effect gives the substrate infinite thermal conductivity and predicts the overgrowth temperature on the basis of the balance between heat generation and radiative heat loss is thus an excellent approximation to the exact situation prevailing when deposition is onto substrates of thickness 1 mm or less. In contrast to this, the fact that the far surface of a substrate, of thickness 1 cm, remains at effectively the initial temperature during the generation of the amount of heat equivalent to the deposition of 10 000 Å of metal at a rate of 1000 Å/sec shows that the semi-infinite substrate approximation is valid for the conditions pertaining to the technological process of lens blooming and mirror coating, etc.

Table 2. Surface temperatures attained by substrates of various thicknesses at various times when heat is generated onto the surface $x=l$ at a rate of 0·2 cal/cm²sec. Thermal diffusivity and conductivity taken as 0·005 and 0·003 respectively

Time t (sec)	l in cm =	1·0	0·1	0·01
1	v_0 °C ·	0·0	2·2	33·3
	v_l	5·3	5·5	33·6
3	v_0	0·0	8·9	100·0
	v_l	5·3	12·2	100·3
6	v_0	0·0	18·9	200·1
	v_l	13·0	22·2	200·4
10	v_0	0 0	32·3	333·6
	v_l	16·8	35·6	333·9

§ 7. Temperature of a Three-Dimensional Aggregate During Growth

The preceding theory is relevant to growth by a continuous film mechanism. In the first stages of overgrowth, however, the experimental indications are that three-dimensional aggregates, rather than two-dimensional monolayers, are formed in many instances. The thermal situation when growth occurs in this way is markedly different from that considered previously and merits separate treatment as the three-dimensional aggregate must approximate to a point source of heat in the initial stages and thus requires a large temperature gradient to allow heat loss from it at a finite rate.

The most obvious model on which to base calculations of the temperature of such an overgrowing aggregate consists of a continuous, growing, disk source of heat which represents the base of the overgrowing aggregate, on a semi-infinite substrate. The temperature at the centre of the disk source would then approximate to the temperature of the aggregate. As the calculation based on this model cannot be handled in terms of tabulated functions, it was found necessary to use the following model instead: The heat generated as the aggregate grows is considered to be released at a point $r=0$ on the surface of semi-infinite substrate. Instead of the aggregate growing on the substrate, however, we consider the closely related hypothetical case in which the aggregate grows into the substrate, equating the volume of the ingrowing aggregate at any instant to the amount of material that has arrived at the 'catchment area' of the aggregate by that time.

The temperature v at a point r in an infinite solid in which heat is continuously liberated at the origin ($r=0$), at a rate Q subsequent to time $t=0$ is given by Carslaw and Jaeger (1948, p. 221) as

$$v=\frac{Q}{4\pi Kr}\operatorname{erf}c\frac{r}{(4\alpha t)^{1,2}}. \qquad \ldots\ldots \quad (6)$$

The equivalent temperature in the hemispherical geometry that our model is based on, i.e. on a semi-infinite, rather than an infinite, substrate will be just twice as large assuming no heat loss from the exposed surface.

When the aggregate grows as a hemisphere with material arriving at a constant rate A cm^3/sec, its volume V will be given by

$$V=\tfrac{2}{3}\pi r^3=At. \qquad \ldots\ldots\ldots \quad (7)$$

The average temperature of the aggregate can be calculated from (6) and is given by

$$v=(2\pi r^3/3)^{-1}\int_0^r (Q/2\pi Kr)2\pi r^2\operatorname{erf}c\left\{\frac{r}{(4\alpha t)^{1/2}}\right\}dr$$

$$=\frac{6Q\alpha t}{\pi r^3 K}\left\{\tfrac{1}{2}\operatorname{erf}\left\{\frac{r}{(4\alpha t)^{1/2}}\right\}+i^2\operatorname{erf}c\left\{\frac{r}{(4\alpha t)^{1/2}}\right\}-0{\cdot}25\right\}.$$

Substituting for r from (7) this becomes

$$v=\frac{4\alpha Q}{KA}\left\{\tfrac{1}{2}\operatorname{erf}\frac{(3At/2\pi)^{1\,3}}{(4\alpha t)^{1\,2}}+i^2\operatorname{erf}c\frac{(3At/2\pi)^{1,3}}{(4\alpha t)^{1/2}}-0{\cdot}25\right\}.$$

The highest values of Q are such that $Q\cong 10^4 A$ cal/sec and, as before, $0{\cdot}6\alpha\cong K$, thus giving

$$v=6{\cdot}7\times 10^4\left\{\tfrac{1}{2}\operatorname{erf}\frac{(3At/2\pi)^{1,3}}{(A\alpha t)^{1/2}}+i^2\operatorname{erf}c\frac{(3At/2\pi)^{1/3}}{(4\alpha t)^{1/2}}-0{\cdot}25\right\}. \qquad (8)$$

Equation (8) has been evaluated for various values of the time (t), with $\alpha=0{\cdot}005$ and giving A the realistic value of 10^{-13} cm^3/sec which corresponds to nuclei $1\,\mu$ apart when the mean deposition rate is 1000 Å/sec or $10\,\mu$ apart when the deposition rate is 10 Å/sec. The results of the calculation show that the mean temperature of the aggregate falls to 86°C above that of the substrate after 10^{-6} sec and to 9°C above the substrate temperature after 1 sec.

In actual growth the supply of material to a nucleus would equal the volume rate of deposition of material onto the area eventually supplying the growing aggregate only after the initial transient needed to establish the flow pattern to the nucleus had passed. The temperatures calculated from (8) for very small values of the time are thus too high as instantaneous aggregation of all the material arriving in the 'catchment area' is assumed in the calculation. Neglect of the thermal capacity of the aggregate and the fact that part of the heat of vaporization is lost to the substrate as heat of adsorption before aggregation also causes over-estimation of the temperature but these factors are compensated, at least partially, by the

fact that in the model adopted, heat loss to the substrate takes place through twice the area that it would be through when the aggregate grows on the substrate with a contact angle of 90°.

It can be concluded thus that growth by a three-dimensional aggregation process in which the contact angle is less than 90° can lead to high aggregation temperatures only in the earliest stages of growth when both deposition and surface migration rates are very high. It is not to be expected that the contact angle would be much larger than 90° for the materials commonly encountered in the 'overgrowth by vacuum sublimation' literature, as this would imply a large measure of re-evaporation during the migration to the nucleus process which is in turn incompatible with common experience.

§ 8. Summarized Conclusions

The temperature of an overgrowing film will be essentially that of the substrate when growth is by a continuous film mechanism. High film temperatures thus necessitate low substrate thermal capacities in order that the heat of sublimation released on formation of the film may elevate the temperature of the system appreciably. The possibility of high transient temperatures being attained in the initial stages of growth by a three-dimensional aggregation process exists when deposition is at a very high rate onto a substrate of low thermal diffusivity, and migration of the incident material is rapid. It has been shown in the extreme case, for which calculations were made, that the aggregate temperature drops to less than 100°C above substrate temperature after only 10^{-6} sec and after 1 sec is only 9°C above the substrate temperature.

Acknowledgment

The author would like to express his gratitude to Professor J. H. van der Merwe of the University of Pretoria for his interest in this work.

References

Carslaw, H. S., and Jaeger, J. C., 1948, *Conduction of Heat in Solids*.
Hoffman, R. H., Daniels, R. D., and Crittenden, E. C., 1954, *Proc. phys. Soc. Lond.* B, **67**, 497.
Macey, H. H., 1940, *Proc. phys. Soc. Lond.*, **52**, 625.
Pashley, D. W., 1956, *Advanc. Phys.*, **5**, 173.
Wilman, H., 1955, *Proc. phys. Soc. Lond.* B, **68**, 474.

The Annealing of Electron Irradiation Damage in Graphite†

By W. N. Reynolds and P. R. Goggin
Atomic Energy Research Establishment, Harwell, Berks

[Received May 26, 1960]

Abstract

Samples of pure artificial graphite have been irradiated with electrons at 79°K, and a study made of the annealing characteristics of the changes in electrical conductivity, Hall effect and magneto-resistance. All three quantities show an anomalous annealing range below 120°K, and a simplified band theory calculation indicates that the concentration of both electrons and positive holes increase during irradiation and also in the anomalous annealing range.

§ 1. Introduction

Work described by Austerman and Hove (1955) and in more detail by Austerman (1958) indicated the importance of resistivity annealing studies in irradiated graphite from about 80°K upwards. These authors exposed samples of graphite to 1·25 Mev electron beams at liquid helium temperatures and then observed the effects on resistivity at 4°K of successive 1 min anneals to temperatures up to about 280°K—the so-called pulse annealing technique. The dose they used, about 20–40 microampere-hours, was such as to cause an initial resistivity increase of about 5%. In the annealing range up to about 80°K, they found only very small resistivity changes, but from 80°K to 110°K a further sharp increase in resistivity was observed amounting to about 30–40% of the initial increase on irradiation. At higher temperatures the behaviour was more normal, the resistivity tending monotonically at a decreasing rate to the unirradiated value. Austerman suggested that the anomalous resistivity increase could be due either to a release of trapped charge carriers or to an increase in density of carrier scattering centres and that experiments could be carried out to separate these factors. He also pointed out that irradiations in liquid nitrogen at 77°K would be suitable, provided that that temperature was held to within a few degrees.

It may be doubted whether the two factors mentioned by Austerman can be separated so easily. Thus, in the work described by Eatherly (1953)‡ on the annealing of pile-irradiated graphite, an anomolous resistivity increase observed on pulse-annealing up to 1300°C was accompanied

† Communicated by the Authors.

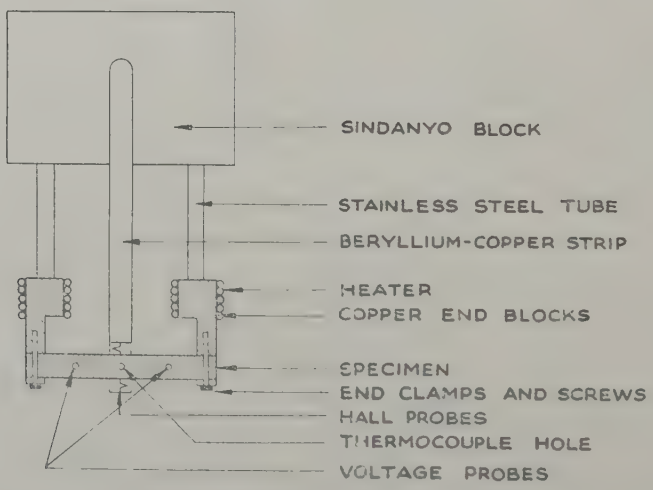
‡ The authors are indebted to the Director of the USAEC Division of Classification for permission to mention some of Dr. Eatherly's results, which are only available in a classified report.

by anomalies in the magneto-resistance and Hall-effect curves. However, the interpretation of such results in terms of band theory parameters such as carrier density and mobility is now on a firmer basis, owing to recent theoretical work. The foundation of this work was laid by Slonczewski and Weiss (1958) and by McClure (1957) who generalized previous band calculations by the use of group theory. The experimental results of Galt *et al.* (1956) as analysed by Nozières (1958) on cyclotron resonance and those of Soule (1958 a) as analysed by McClure (1958) and Soule (1958 b) on the oscillatory galvanomagnetic effects in graphite have permitted elaboration of the band structure model second only to that of the commoner semiconductors.

§ 2. Experimental Method

Pulse annealing studies have been made on samples of artificial graphite irradiated by means of a 4 Mev electron linear accelerator at liquid nitrogen temperature. The properties studied were resistivity, Hall effect and magneto-resistivity at a field of 5600 gauss. The specimens were made of Acheson grade A pure pile graphite, cut parallel to the direction of extrusion,

Fig. 1



Apparatus for the electron irradiation and thermal annealing of graphite rods.

in rods 3 mm square and 25 mm long. A central 36 s.w.g. hole was drilled perpendicular to the length to take a thermocouple, made of 40 s.w.g. chromel–alumel wire, as shown in fig. 1. The voltage contacts were inserted into holes at 6 mm on either side of the centre, defining the useful length as 12 mm. The Hall contacts were fine beryllium–copper wire springs, and these were supported, together with the copper end-clamps for current supply on a block of Sindanyo. The whole device was mounted on 18 in.

long stainless steel tubes for successive insertion into the accelerator head and the annealing vacuum chamber, and was completely wired before irradiation, the pre-irradiation values being determined under similar conditions to those for the post-irradiation and annealed values.

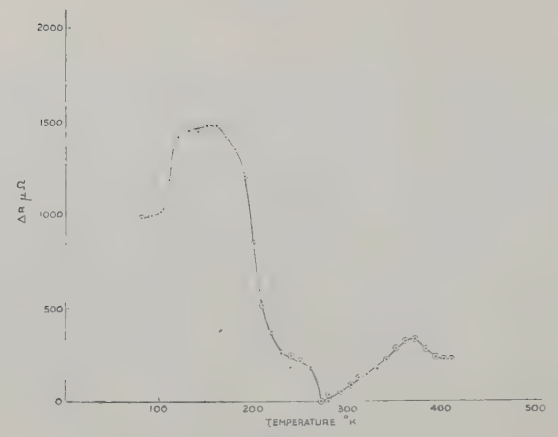
The irradiation chamber was a brass box screwed to the accelerator. The beam was admitted through a thin copper window and a slit the same size as the specimen in the brass wall was used as a stop. An exit window was also provided. The beam current into the specimen was in the range 10–30 μamp, and exposures of 1–2 hours produced resistivity increases of 5–10%. The irradiation chamber was completely filled with liquid nitrogen supplied by a nitrogen pressurized siphon from a Dewar vessel, at a sufficient rate to keep it flowing out of a hole at the top in liquid form. It was found advisable to include a piece of copper gauze to reduce the risk of troublesome minor explosions, the exact cause of which was not determined. Continuous readings of both temperature and resistance were made during the runs to ensure that the temperature rise in the specimens was less than 2°C. The resistance rose linearly with the exposure time, but the change on switching the beam on or off was very small. At the end of the irradiation period, the specimen end of the apparatus was immersed in liquid nitrogen, and then transferred to the annealing chamber.

The annealing chamber was a glass vacuum system which could be pumped to better than 10^{-5} mm Hg, and was immersed in a special Dewar flask containing liquid nitrogen between the poles of an electromagnet. Heat for the annealing programme was supplied by small electrical heaters wound on the copper end clamps. The current supply for the electromagnet was carefully stabilized. The thermocouple and electrical leads were brought out through special bushes. The directions of specimen and magnet current were separately reversed for each reading. The annealing chamber was filled with pure dry nitrogen or helium gas during readings, as this was found to give more stability to the base temperature used for the pulse annealing, viz. 80°K. No difference was observed in specimen properties between readings in nitrogen and those in helium.

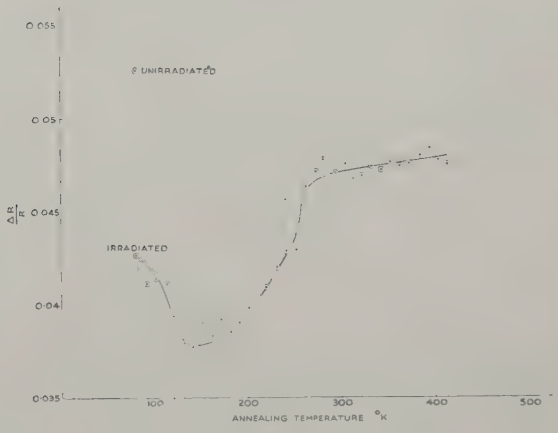
§ 3. Results

Figures 2 (*a*), (*b*) and (*c*) show the behaviour of the resistivity change, magneto-resistivity and Hall coefficient as a function of annealing temperature at about 10°C intervals. The resistivity curve, fig. 2 (*a*) shows the same general trends as those of Austerman's curves, as well as some additional features. The resistivity rise begins at 80°K and is almost completed at 120°K. However, a further slight increase is now seen up to 160°K, at which temperature a rather sharp fall off is indicated. The fall in resistivity from 160°K to 270°K is not so simple as suggested by Austerman, and the original irradiated value is crossed slightly below, rather than above 200°K. All our curves then show a well-marked minimum at 270°K–280°K, although it is not certain that the excess resistivity falls so low in every case because, at any rate in the earlier results, we took no special care to

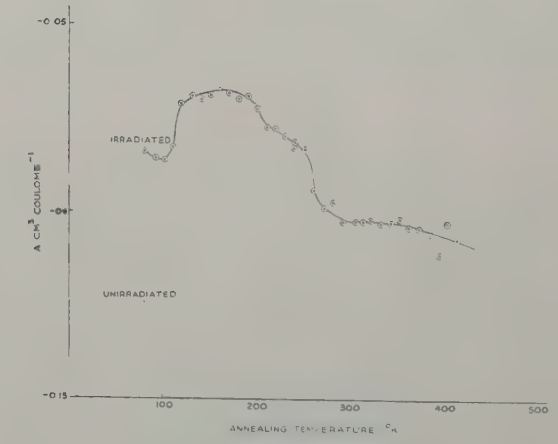
Fig. 2



(*a*)



(*b*)



(*c*)

(*a*) Excess resistance ΔR due to irradiation and annealing as a function of annealing temperature. (*b*) Relative increase in specimen resistance $\Delta R/R$ in a magnetic field of 5600 gauss on irradiation and annealing as a function of annealing temperature. (*c*) Hall coefficient A in cm^3/Coulomb on irradiation and annealing as a function of annealing temperature.

study this unexpected feature. All specimens show a second maximum around 400°K, although not generally as sharp as that given here.

The magneto-resistance and Hall-effect curves in figs. 2 (*b*) and (*c*) are somewhat unexpected in that they both indicate an apparent increase of the irradiation damage effects for anneals in the 80°K to 120°K range. More straightforward annealing again commences at about 160°K, but there is no extreme value at 270°K showing that the resistance minimum is due to a fortuitous balance of the carrier densities and mobilities; its magnitude may depend critically on the local structure. In any case, the resistivity when measured alone now appears a rather unreliable guide to radiation damage annealing.

Figure 3 shows the experimental results normalized to the initial changes on irradiation. All three show the initial rise of 40–50%, but the resistivity falls away more rapidly on the high temperature side. The particular specimen illustrated was irradiated at a beam current of 14 μamp for about 105 min. It is noteworthy that specimens given about twice this dose gave normalized resistivity and magneto-resistivity curves closely agreeing with those of fig. 3, but that the normalized Hall effect showed an initial rise of only about 20%.

§ 4. Discussion

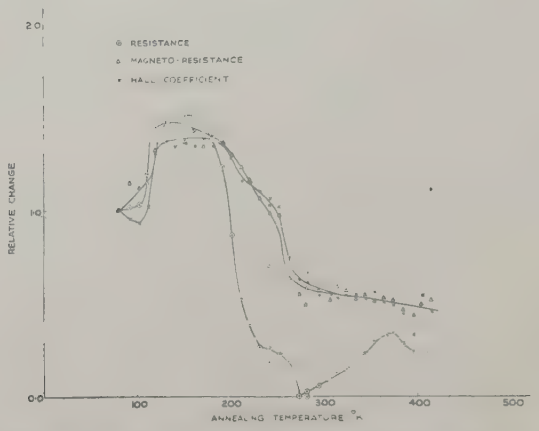
We note first that the theory of Kinchin and Pease (1955) gives the mean number of atoms displaced per primary knock-on for 4 MeV electrons in graphite as 2·8, whereas that for 1·25 MeV electrons as used by Austerman is only slightly greater than unity. Thus, in our specimens, the damage may be more complex. However, many primary displacements must still be in the form of simple interstitial-vacancy pairs, and in view of the correspondence in the resistivity behaviour, we shall not consider these complications at present. The Kinchin and Pease theory also indicates a total density of $1{\cdot}2 \times 10^{16}$ displaced atoms/cm^3 for this specimen.

Many people have observed changes of graphite properties similar to those shown here on initial irradiation. In general, they are explained (see e.g. Hennig and Hove 1955) by supposing that the defects trap electrons, thus lowering the Fermi level and making the conduction tend from n to p-type, and also, since the charged defects act as scattering centres, decreasing carrier mobilities.

The similar sharp increase in the 80°–120°K temperature range in the normalized values of the quantities measured, as shown in fig. 3 indicates that some re-arrangement of the original damage is taking place, quite distinct from direct re-integration. This could be a dispersion of interstitial-vacancy pairs, if it is supposed that, over a certain range, there is a repulsive force between the two components, as has been suggested, for example, by Wertheim (1959) in the case of silicon. In order to account for the decreasing magneto-resistance, it would then be necessary to postulate that this dispersion permitted further electron trapping, because we know (see e.g. Mott and Jones 1936) that the scattering power of a

defect is proportional to the square of its change. Thus dispersion without additional trapping would be expected to give less scattering and higher mobility and magneto-resistance, but further electron trapping would account for the observed changes in both magneto-resistance and Hall effect.

Fig. 3



Changes in resistance, magneto-resistance and Hall coefficient normalized to to initial changes on irradiation as a function of annealing temperature.

This hypothesis may be examined in greater detail. Using unpublished expressions derived by Johnston (given in the appendix) we have calculated values of the mobilities μ_+ and μ_- and effective densities n_+ and n_- of holes and electrons respectively, corresponding to the data in fig. 2. These expressions are based on an overlapping two-band cylindrical model of graphite, using the theory of Jones and Zener (1934). A more complex model based on single crystal studies would not be appropriate in view of the uncertainties over corrections for orientation, density and crystallite boundary effects in our materials. The Jones-Zener theory solves the Boltzmann equation in terms of an expansion in powers of μB, where B is the applied magnetic field, which applies to our results since $\mu B \sim 0{\cdot}1$. In the case of Soule's (1958a) work on single crystals, where $\mu B > 1$, the form of solution given by McClure (1956) is required.

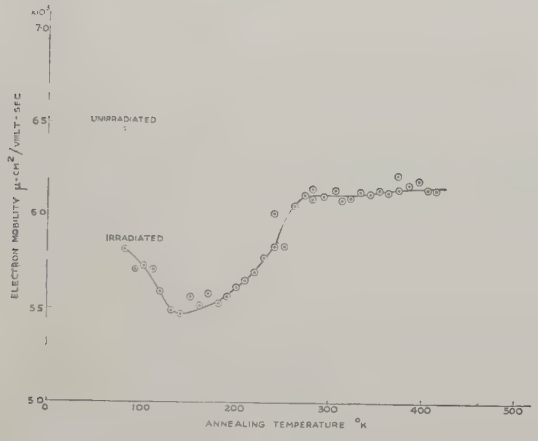
If the relaxation times for the two types of current carrier are assumed equal, the calculations of Johnston (1955) suggest that

$$\frac{\mu_+}{\mu_-} = b \sim 0{\cdot}8$$

so that for this value of the mobility ratio we may calculate the values of μ_+, μ_-, n_+ and n_- from the three corresponding experimental values of A, the Hall coefficient, ρ the resistivity and $\Delta\rho/\rho$ the magneto-resistance observed on application of the standard field B. Figures 4 and 5 give the values obtained for μ_- as a function of annealing temperature (μ_+ is merely 0·8 of this value at any point) and n_+ and n_- respectively. The mobilities and carrier densities were all lower than those obtained from

the single-crystal work at the same temperature. Crystallite boundary scattering is thought to be important in artificial graphite, and to account for the opposite sign of the temperature coefficient of resistance by limiting the carrier mean free path. This would also account for the low mobilities observed, but the low carrier densities are less easily explained.

Fig. 4



Electron mobility as a function of annealing temperature.

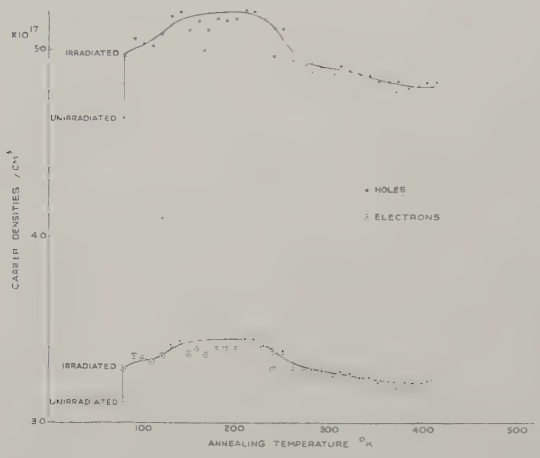
The most striking feature of fig. 5, however, is the increase of concentration found for both types of carrier both on irradiation and in the lower part of the annealing range. Such behaviour is not compatible with the simple two-band model and constant mobility ratio which we have used, and may be connected with the type of band distortion described by Parmenter (1955) which could cause an increase in the band overlap in graphite.

It is observed in the appendix that for the model we have used, the difference between hole and electron concentrations at any point on the annealing curve must be equal to the density of trapped electrons. Carrier scattering is presumably increased by trapped electrons, and it is therefore of interest to study the relationship between n_+ and n_- and the carrier mobilities. Figure 6 shows the value of $n_+ - n_-$ as a function of $1/\mu_-$ for an assumed value of $b = 0{\cdot}8$. It is interesting that the change in trapped electron density on irradiation is about $1{\cdot}5 \times 10^{16}/\text{cm}^3$—quite close to the calculated defect density, and the two could in fact be made equal by putting $b \sim 0{\cdot}85$. The increase in trapped carrier density in the first annealing phase would then require that some defects acquire more than one charge. The loop in the curve for later phases of annealing shows that processes other than direct re-integration are involved.

Other workers have noticed the rapid change in magneto-resistance compared with resistance for pile irradiation, which leads to the above

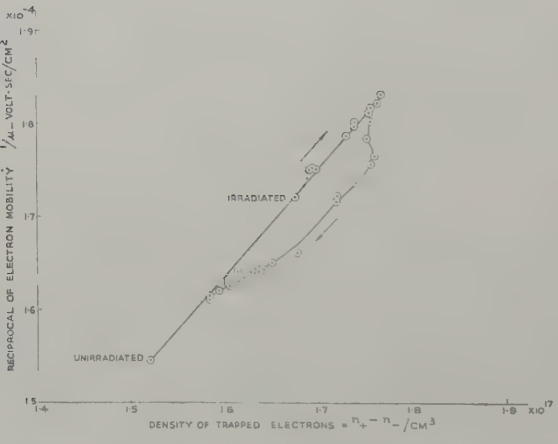
conclusions, notably Kinchin (1954) but the only detailed work seems to be that of Eatherly (1953). He studied the resistivity, magneto-resistivity and Hall effect on AWG graphite specimens cut parallel to the axis of extrusion and irradiated up to a dose of 460 MWD. An analysis of his

Fig. 5



Carrier densities as a function of annealing temperature.

Fig. 6



Variation of density of trapped electrons $n_t = n_+ - n_-$ with electron mobility reciprocal $1/\mu_-$.

results similar to that which we have used shows that the density of holes increases monotonically and the carrier mobilities decrease monotonically with dose, but that the electron density rises to a maximum at about 3 MWD and then falls off.

§ 5. Conclusions

The extension of Austerman's work by simultaneous measurements of magneto-resistance and Hall effect has given results of considerable interest. The analysis of these results shows that the simple band-theory model cannot provide a full explanation.

Other annealing experiments, such as the study of the region around 0°c now appear to be of interest, and it would also be profitable to compare higher dose electron damage with lower dose neutron damage. However, detailed atomic mechanisms are not yet identifiable.

Acknowledgments

The authors would like to express their thanks to their colleagues in the Theoretical Physics Division for discussion of the results of this work and for the performance of computations. They are particularly indebted to Dr. D. F. Johnston who derived a new approximation for the magneto-resistance relationship which he describes in the appendix.

APPENDIX

By D. F. Johnston

Galvanomagnetic Formulae for an Overlapping Cylindrical Two-band Model of Graphite

Hall coefficient

$$A=\frac{E_y}{B_z I_x}=\frac{1}{e}\frac{(\mu_+^2 n_+-\mu_-^2 n_-)}{(\mu_+ n_+ +\mu_- n_-)^2}\cdot\frac{1}{\cos^2\theta}\,.$$

Resistivity

$$\rho=\frac{E_x}{I_x}=\frac{1}{e(\mu_+ n_+ +\mu_- n_-)}\,.$$

Magneto-resistance

$$\frac{\Delta\rho}{\rho}=\frac{B^2\cos^2\theta}{\mu_- n_- +\mu_- n_-}\left[\mu_+^3 n_+ +\mu_-^3 n_- -\frac{(\mu_+^2 n_+ -\mu_-^2 n_-)^2}{\mu_+ n_+ +\mu_- n_-}\right]$$

where n_+ and μ_+ denote the concentration (n/cm^3) and mobility (cm^2/sec volt) respectively of the 'holes' in the lower band. Similarly n_- and μ_- denote these features for electrons in the upper band.

The magnetic field B_z is in gauss divided by 10^{-8}.

The electronic charge e is in coulombs, i.e. $e=1{\cdot}6\times10^{-19}$.

These formulae are valid only if μ_+, μ_- and B are such that $(\mu_+ B_z)^2$ and $(\mu_- B_z)^2$ are negligible relative to unity.

These formulae are derived for the case where the axis of the cylindrical energy surfaces is contained in the y–z plane at an angle θ to the z-axis. The magnetic field is such that $B_x=0=B_z$.

The energy surfaces of the model are of the form

$$\epsilon_+=\epsilon_0-\lambda_+(k_1^2+k_2^2),$$
$$\epsilon_-=\lambda_-(k_1^2+k_2^2),$$

where ϵ_0 is the 'overlap' of the two bands. We define the mobilities of 'holes' and electrons in the form

$$\mu_+ = \frac{2e\tau_+\lambda_+}{h^2}, \quad \mu_- = \frac{2e\tau_-\lambda_-}{h^2}$$

where τ_+ and τ_- are relaxation times for 'holes' and electrons respectively. The formulae have been simplified by neglecting the variation of τ_+ and τ_- near the Fermi surface.

The concentrations of 'holes' and electrons, n_+ and n_-, are defined by

$$n_+ = \frac{4\pi kT}{c\lambda_+} \ln\{1+\exp(\epsilon_0-\eta)/kT\},$$

$$n_- = \frac{4\pi kT}{c\lambda_-} \ln\{1+\exp\eta/kT\},$$

where c denotes the c-spacing in in graphite and η denotes the chemical potential. With this definition:

$$n_+ - n_- = \text{concentration of trapped electrons.}$$

Orientation and Density Corrections

We have assumed that in our polycrystalline sample of graphite, the crystallites are orientated so that their c-axes are contained in the y–z plane. Accordingly we replace $\cos^2\theta$ by $\frac{1}{2}$ in the formulae given in (1).

If α is the ratio of the observed density to the theoretical density of graphite, the true current density I_x is related to the apparent current density I_x' by the relation $I_x = I_x'/\alpha$. We have taken $\alpha = 0{\cdot}8$

References

Austerman, S. B., 1958, Atomics International Report No. NAA-SR-2457.
Austerman, S. B., and Hove, J. E., 1955, *Phys. Rev.*, **100,** 1214.
Eatherly, W. P., 1953, Classified USAEC Report.
Galt, J. K., Yager, W. A., and Dail, H. W., 1956, *Phys. Rev.*, **103,** 1586.
Hennig, G. R., and Hove, J. E., 1955, First Geneva Conference Paper 7P/666.
Johnston, D. F., 1955, *Proc. roy. Soc.* A, **227,** 349.
Jones, H., and Zener, C., 1934, *Proc. roy. Soc.* A, **145,** 269.
Kinchin, G. H., 1954, *J. nucl. Energy,* **1,** 124.
Kinchin, G. H., and Pease, R. S., 1955, *Rep. Progr. Phys.*, **18,** 1.
McClure, J. W., 1956, *Phys. Rev.*, **101,** 1642 ; 1957, *Ibid.*, **108,** 612 ; 1958, *Ibid.*, **112,** 715.
Mott, N. F., and Jones, H., 1936, *The Theory of the Properties of Metals and Alloys* (Oxford : University Press), p. 292.
Nozières, P., 1958, *Phys. Rev.*, **109,** 1510.
Parmenter, R. H., 1955, *Phys. Rev.*, **97,** 587.
Slonczewski, J. C., and Weiss, P. R., 1958, *Phys. Rev.*, **109,** 272.
Soule, D. E., 1958 a, *Phys. Rev.*, **112,** 698 ; 1958 b, *Ibid.*, **112,** 708.
Wertheim, G. K., 1959, *Phys. Rev.*, **115,** 568.

The Phonon Drag Component of the Thermoelectric Power of the Alkali Metals at Low Temperatures†

By M. Bailyn
Northwestern University, Evanston, Illinois, U.S.A.

[Received September 6, 1959 ; and in revised form March 4, 1960]

Abstract

The theory of phonon drag in the thermoelectric power is developed from an equation derived in a previous paper. It is shown that important in the understanding of the phenomena are the relative probabilities $\alpha(j\mathbf{q}; \mathbf{q}+\mathbf{K})$ that a phonon of wave vector $\mathbf{q}$ and polarization j will interact with electrons via a process of umklapp type $\mathbf{K}$, where $\mathbf{K}$ is a reciprocal lattice vector or zero. $\alpha(j\mathbf{q}, \mathbf{q}+\mathbf{K})$ is the probability relative to all the processes into which the jq phonons can enter, including phonon–phonon collisions and phonon-impurity interactions. It is shown by means of a ' baby spectrum ' that the use of α results in the domination of the phonon drag effect by the transverse-like phonon modes, and in the watering down of any ' bulge ' effect of the Fermi surface. Rough calculations are made that show that one can get qualitative agreement with experiment over at least part of the temperature range from 2 to 20°K. It is shown that with neglect of phonon–phonon and phonon-impurity interactions the phonon drag part S_g of the thermoelectric power is the sum of two terms, a normal (negative) term S_g^-, which is exactly what would be calculated if all processes were non-umklapp processes, plus a term S_g^+ which is anomalous (positive) in sign and is the effect on the electrons of the Bragg reflections abstracted from the umklapp processes. A simple derivation of the basic equation is given.

§ 1. Introduction

In 1958, MacDonald *et al.* measured the thermoelectric power of the alkali metals down to 2°K. They found that the magnitudes of this quantity were in all cases many times that predicted by the normal theory, in caesium as much as about 30 times larger, in sodium about 10 times larger. They also found that the sign becomes anomalous, namely positive, in caesium and rubidium, and that an oscillation takes place in potassium between 2 and 10°K. These authors showed that the phonon-drag effect might possibly explain the measured magnitudes since it had been shown by MacDonald (1954) (without the factor $\frac{1}{3}$) and Hanna and Sondheimer (1957) that on neglect of phonon–phonon interactions the phonon-drag contribution to the thermoelectric power is given by

$$S_g = -\frac{1}{3}\frac{C}{N|e|} = -\frac{77{\cdot}8\kappa}{|e|}\left(\frac{T}{\Theta}\right)^3 \quad \ldots . T \ll \Theta, \quad . \quad . \quad . \quad (1.1)$$

where C is the specific heat of the lattice, and where N is the number of

† Communicated by the Author. This work was partially supported by a National Science Foundation Grant (U.S.A.).

atoms per unit volume. Θ is a Debye temperature. Such a term does give large magnitudes in the temperature range considered. But it was thought that this would always be a negative contribution and hence there was no explanation of the change in sign to positive in Rb and Cs, nor of the oscillation in K.

In 1958, the author developed a theory to take into account the effect of the non-equilibrium component of the phonon distribution by means of a Variation Principle. The transport coefficients were derived, and in particular the phonon-drag term S_g in the thermoelectric power was formulated. It was found that the umklapp (UK) processes in the scattering of electrons (in monovalent metals) almost always yielded an anomalous or positive contribution to S_g, whereas the normal, non-UK processes always yielded a normal (negative) sign in S_g. It was therefore suggested that the UK contribution to the phonon-drag effect would be able to account for the anomalous sign as well as the anomalous magnitude of the measured thermoelectric power. It was cautioned however that the actual effect was the result of a 'delicate balance' between very large positive and negative contributions, and that an extremely complicated computation would be required to see if the theory in fact matched the experiment. In addition, not the ordinary electron–phonon matrix element appears, but rather one normalized in a different fashion, and consequently quite a bit more difficult to know how to handle.

There have been other approaches to the same problem. Notably, Hanna and Sondheimer (1957) considered the solution of the Boltzmann equation taking into account the non-equilibrium component of the phonons. Their expressions would correspond to a first approximation to the solution obtained by the author were it not for the fact that they did not take into account umklapp processes. For other references see Bailyn (1958).

Ziman (1959) had however developed a theory which is also a Variational Principle and which does take into account umklapp processes. By ignoring phonon–phonon processes, he was able to compute the phonon-drag term in the thermoelectric power, and found that by postulating a large effect arising from the distortion of the Fermi surface, he was able to get tolerable agreement between theory and experiment.

We have rather serious objections to both the theory and the computations of Ziman, and the purpose of this paper is to develop the theory according to the lines we originally suggested. In § 8 we compare the two approaches.

In § 2 we derive the basic expression of Bailyn (1958) in a very simple manner. This derivation evolved out of conversations with Dr. D. K. C. MacDonald. The confidence that a simple derivation was possible was entirely his. In § 3 we separate S_g into the intrinsically positive portion (S_g^+) and the intrinsically negative portion (S_g^-). In § 4 we derive an approximate expression for S_g^+ from which calculations are subsequently made. In §§ 5 and 6 we construct and discuss the phonon spectrum from

which we draw a qualitative picture of the nature of the contributions to S_g^+.

§ 2. Simple Derivation of the Basic Equation (With D. K. C. MacDonald)†

We regard the phonons as proceeding down a metal bar from one at temperature T_1 to the other at T_2. The length of the bar is δx. We consider a quantity $\delta N(j\mathbf{q})$ which is the net number of phonons of wave vector $\mathbf{q}$ and polarization j in the bar at any instant that must be annihilated sooner or later before reaching the cold end of the bar in order to maintain the boundary conditions of temperature. (Going in the $-x$ direction, the same number of $-q, j$ phonons must be created. q is supposed to have a positive x-component, so that $-q$ has a negative x-component.) To first order in the non-equilibrium component of the distribution, this is given by

$$\delta N(j\mathbf{q}) = \frac{\partial N_0(j\mathbf{q})}{\partial x}\,\delta x, \qquad \ldots\ldots\ (2.1)$$

where $N(j\mathbf{q})$ is the equilibrium number of phonons at T_1 or T_2. (It does not matter for our purpose where in the interval T_1, T_2 we evaluate $\partial N_0/\partial x$.)

Let us now define $\alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K})$ to be the relative probability that any one of these excess phonons will be destroyed (i.e. will interact) by means of an UK process of type $\mathbf{K}$. A given $\mathbf{q}j$ phonon may get destroyed by (1) normal interactions, with electrons denoted by $\mathbf{K}=\mathbf{K}_0=0$, (2) umklapp interactions with electrons, denoted by $\mathbf{K}_n$ for the nth reciprocal lattice vector (a given phonon may possibly interact by means of more than one type of UK process), and (3), all the other processes not involving electrons, such as phonon–phonon interactions, boundary scattering impurity, etc. The processes in the last category are described by a probability B_0 which is the sum of the probabilities for the separate processes. The quantity α will be taken from Bailyn (1958, eqn. (69)). (See also eqns. (11) and (13) and the dagger footnote page 1593 of that reference.)

$$\alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K}) = \frac{B(j\mathbf{q};\ \mathbf{q}+\mathbf{K})}{B_0 + \sum_{\mathbf{K}'}^{[j\mathbf{q}]} B(j\mathbf{q};\mathbf{q}+\mathbf{K}')}. \qquad \ldots\ldots\ (2.2)$$

The B's represent 'absolute probabilities'. They are given by Bailyn (1958, eqn. (65)) and are met again in (5.5) below. The $\mathbf{K}'$ sum in (2.2) means a sum over all the umklapp processes which a given phonon $\mathbf{q}j$ can enter into. There is a temperature dependence in α arising from the fact that B_0 is temperature dependent, getting in fact much larger as T gets larger. At all but the lowest temperatures, B_0 is so large that the α's become small enough to make phonon drag effects negligible.

† The opinions presented in the remainder of this paper are not necessarily shared by Dr. MacDonald.

To calculate the thermoelectric power, we seek the effective x-component of the force, F_x, on the electrons resulting from phonons relaxing by mechanisms (1) and (2) above. Each collision of type K_n will transfer a crystal momentum

$$P_{\mathbf{q},\mathbf{K}} = \hbar(\mathbf{q}+\mathbf{K}) \qquad . \; . \; . \; . \; . \; . \; . \; . \; (2.3)$$

into the electron system. Th probable x-component of the momentum introduced into the electron system by the *excess* $\mathbf{q}j$ phonons relaxing by $\mathbf{K}_n$-type collisions is then

$$\delta N(j\mathbf{q})\alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K}_n)[\mathbf{P}_{\mathbf{q},\mathbf{K}}]_x. \qquad . \; . \; . \; . \; . \; . \; (2.4)$$

We have to use the relative probabilities α, because the number $\delta N(j\mathbf{q})$ of phonons *must* be destroyed somehow or other in order to maintain the temperature boundary conditions. Therefore the sum of probabilities for all the interactions that these excess $\mathbf{q}j$ phonons can enter into must be 1. This 'renormalization' is characteristic of all phonon–drag calculations.

A sum over n (including the normal term $n=0$) would then give the total momentum introduced by the $\mathbf{q}j$ phonons. The time it would take would be the time $\delta t(j\mathbf{q})$ taken by a $\mathbf{q}j$ phonon which managed to travel from one end of the bar to the other before being destroyed. Thus the force on the electrons from the $\mathbf{q}j$ phonons is equal to the rate of change of momentum caused by phonon destruction and is

$$F_x(j\mathbf{q}) = \sum_{\mathbf{K}}^{[j\mathbf{q}]} \frac{\partial N_0}{\partial x} \frac{\delta x}{\delta t(j\mathbf{q})} [P_{\mathbf{q},\mathbf{K}}]_x \alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K}). \qquad . \; . \; (2.5)$$

But $\delta x/\delta t(j\mathbf{q})$ is just the x-component of the drift (group) velocity $\nabla_q\omega(j\mathbf{q})$ of the $j\mathbf{q}$ phonons where $\omega(j\mathbf{q})$ is the frequency of these phonons. Further the force F_x may be set equal to $-|e|\mathscr{E}_x N$, where N is the number of electrons in the bar, and $\mathscr{E}_x$ is the average electric field that would produce this force. In the usual sense, an electric field is the same on every electron, but the force on different electrons would be different according to (2.4). We thus have here an average force (which would be the measured force). Finally we write

$$\frac{\partial N_0}{\partial x} = \frac{z}{(e^z-1)(1-e^{-z})} \frac{1}{T}\frac{dT}{dx}, \qquad . \; . \; . \; . \; . \; . \; (2.6)$$

where

$$z = \frac{\hbar q v(j\mathbf{q})}{\kappa T}. \qquad . \; . \; . \; . \; . \; . \; . \; . \; . \; (2.7)$$

We have used the fact that the frequency $\omega(j\mathbf{q})$ of a phonon is $qv(j\mathbf{q})$ where $v(j\mathbf{q})$ is the phase velocity. The phonon-drag component of the thermoelectric power is just (2.5) divided by the temperature gradient:

$$S_g = \mathscr{E}_x/(dT/dx) = -(3Ne)^{-1} \sum_{j\mathbf{q}} \sum_{\mathbf{K}}^{[j\mathbf{q}]} \alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K})\hbar(\mathbf{q}+\mathbf{K}) \,.\, \nabla_q\omega\partial N_0/\partial T.$$

$$. \; . \; . \; (2.8)$$

We have contemplated a cubic crystal so that $P_x v_x$ was replaced by $(1/3)\mathbf{P}\,.\,\mathbf{v}$.

This expression for free electrons follows from the γ_0/α_0 term of eqn. (86) of our 1958 article, when eqns. (79) and (77) are used. (For the extension to non-spherical energy surfaces and several bands, see Bailyn (1960), Appendix H.)

We see from (2.8) that for an UK process, $\mathbf{q}+\mathbf{K}$ is directed away from $\nabla_q\omega$ (which is generally in a direction close to that of $\mathbf{q}$), and the contribution is positive, i.e. anomalous. Normal processes ($\mathbf{K}=0$) give a normal sign.

§ 3. Simplification Arising from Neglect of Phonon–Phonon Collisions

We first make a preliminary approximation, which is to neglect phonon–phonon and other relaxation processes not involving electrons. This implies

$$\sum_{\mathbf{K}}^{[j\mathbf{q}]} \alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K})=1. \qquad (3.1)$$

We then go back to (2.8) and separate the right-hand side into a $j\mathbf{q}\mathbf{K}$ summation containing the $\mathbf{q}$ of the momentum factor $\mathbf{q}+\mathbf{K}$ plus a $j\mathbf{q}\mathbf{K}$ summation containing the $\mathbf{K}$ of the momentum factor. The term in $\mathbf{q}$ will then contain a sum over $\mathbf{K}$ which is exactly the left-hand side of (3.1) and hence may be let equal to 1. We thus get the separation.

$$S_g = S_g^{+} + S_g^{-}, \qquad (3.2)$$

$$S_g^{-} = -\frac{1}{3N|e|}\sum_{\mathbf{q}}\sum_{j}\hbar\mathbf{q}\,.\,\nabla_q\omega\frac{\partial N_0}{\partial T} = -C/(3Ne), \qquad (3.3a)$$

$$S_g^{+} = -\frac{1}{3N|e|}\sum_{\mathbf{q}}\sum_{j}\sum_{\mathbf{K}\neq 0}^{[j\mathbf{q}]}\alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K})\hbar\mathbf{K}\,.\,\nabla_q\omega\frac{\partial N_0}{\partial T}. \qquad (3.3b)$$

A number of points needs to be made apropos of these equations. First, in (3.3 *a*) C stands for the specific heat

$$C=\frac{\Delta}{8\pi^3}\sum_{j}\int d^3q\hbar\omega\frac{\partial N_0}{\partial T}, \qquad (3.4)$$

(where Δ is the crystal volume); thus the relation in (3.3 *a*) is possible only if we identify

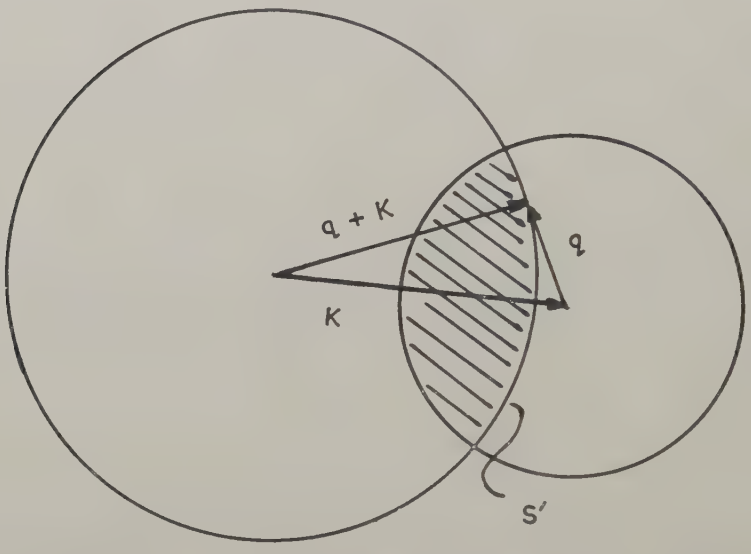
$$\mathbf{q}\,.\,\nabla_q\omega = qv \qquad (3.5)$$

This is not in general true, but it is shown in the Appendix to be the case for the particular spectrum for which we make calculations. This specific heat relation (3.3 *a*), previously obtained by Hanna and Sondheimer (1957), and MacDonald (1954), is thus shown to arise in the more general considerations of this paper, provided phonon–phonon and phonon-impurity interactions are neglected, and provided the spectrum satisfies (3.5).

Second, the sum over $\mathbf{K}$ in (3.3 *b*) is set equal to 12 times the contribution from each $\mathbf{K}$. This is the case because the contribution of each $\mathbf{K}$ is the same, and there are 12 nearest neighbour $\mathbf{K}$'s. The region of $\mathbf{q}$ integration for a given $\mathbf{K}$ is the shaded volume of fig. 1. It is the overlap volume between the q-sphere of radius $q_{\max}$ centred at $q=0$, and the sphere of

radius $2k_0$ (k_0 being the wave vector of the electron at the top of the Fermi sphere) centred at the origin of the reciprocal lattice vector **K**. The tip of **K** is at $q=0$, (i.e. 000) and the tail is at, say, the 011 reciprocal lattice point. In UK interactions an electron wave vector changes from **k** to **k′** with $|\mathbf{k}-\mathbf{k}'|\leqslant 2k_0$ and $\mathbf{k}-\mathbf{k}'=\mathbf{q}+\mathbf{K}$. Hence the permitted q-volume for the $\mathbf{K}=011$ type of umklapp process is the overlap volume as described above.

Fig. 1



The geometry of umklapp processes. The surface S′ is the surface for which table 1 is computed. It is that part of the surface of the shaded volume that is part of the surface of the sphere of radius $|q+K|$ when $|q+K|=2k_0$.

We make the separation into S_g^- and S_g^+ for several reasons. It has the advantage of providing the anomalous sign contribution as an extra term to be added to the 'normal' specific heat term. It has the advantage that, unless details of the electron number N are investigated, many of the complications arising from electron quantities (such as shape of Fermi surface) appear only in S_g^+†, whereas S_g^- in principle can be computed by one of several standard methods, or can be taken from experiment. And finally it has the advantage that the two terms have a definite physical significance. S_g^- is the contribution which would occur if for some reason all the processes were normal. And S_g^+ is the contribution that is provided by the Bragg reflections abstracted from the umklapp interactions. The anomalous sign comes specifically from that part of the umklapp interactions which are the Bragg reflections.

† See Bailyn (1960), Appendix H for more details of these effects.

It should be kept in mind that in (3.3) the phonon–phonon and phonon-impurity interactions have been neglected. When these are taken into account, the relation in (3.3 *a*) is in particular no longer valid, and in general the α's become smaller. One may expect that although the onset (as T rises) of phonon-drag may well be represented by the approximation (3.1), it is quite possible that the dwindling away of the effect may require a more detailed knowledge of the α. This is in fact our explanation of the poor results for caesium above 11°K (see § 7).

§ 4. An Approximate Expression for the S_g^+'s

We shall leave S_g^- as in (3.3 *a*) and assume that the standard calculations or measurements of specific heat will determine it for our purposes. S_g^+ has to be calculated separately however. We do not pretend to any great accuracy in this calculation, but the result should give some indication of the correct temperature dependence. First we write (3.3 *b*) as

$$S_g^+ = -\frac{\Delta}{N|e|}\frac{1}{2\pi^3}\sum_j \mathbf{K}\cdot\int \alpha(j\mathbf{q};\,\mathbf{q}+\mathbf{K})[\nabla_q\hbar\omega]\frac{\partial N_0}{\partial T}d^3q\,. \qquad (4.1)$$

The basic approximation to be made is that for each polarization j the α can be removed at some 'effective value' α_j to be incorporated subsequently in an experimentally determined adjustable parameter. Thus we allow

$$S_g^+ \cong \sum_j A_j' dI_j/dT, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.2)$$

$$I_j = (1/\kappa)\mathbf{K}\cdot\int d^3q[\nabla_q\hbar\omega]N_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.3)$$

$$A_j' = \frac{\Delta}{Ne}\frac{\kappa}{2\pi^3}\alpha_j. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.4)$$

If we use the identity

$$[\nabla_q\hbar\omega]\frac{\partial N_0}{\partial T} = -\frac{\hbar\omega}{T}\nabla_q N_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4.5)$$

then we can form a differential equation for I_j by integrating by parts

$$\left.\begin{aligned}\frac{dI_j}{dT} &= -\frac{\mathbf{K}}{\kappa}\cdot\int d^3q\frac{\hbar\omega}{T}\nabla_q N_0\\ &= I_s + I_j/T,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (4.6)$$

where

$$I_s(T) = -\frac{1}{\kappa T}\mathbf{K}\cdot\int d^3q\nabla_q(\hbar\omega N_0) = -\frac{1}{\kappa T}\mathbf{K}\cdot\int \hat{n}\hbar\omega N_0 dS_q. \qquad (4.7)$$

Here $\hat{n}$ is an outward drawn normal to the surface involved. The total surface would actually be that over the region in which the α's are substantial in magnitude. Part of this surface will always coincide with S′ of fig. 1, and in fact the contribution from this part will be tremendously greater (because of the smaller q's) than from anywhere else on the surface. We then as our second approximation remove the integrand in (4.7) at

an 'effective value' on S′ represented by a kind of Debye temperature Φ_j:

$$I_s(T) \cong S_j \left\langle \frac{\hbar\omega}{\kappa T} N_0 \right\rangle_{S'}, \quad \ldots\ldots\ldots \quad (4.8)$$

$$\left\langle \frac{\hbar\omega}{\kappa T} N_0 \right\rangle_{S'} \cong \frac{\Phi_j}{T} [\exp(\Phi_j/T) - 1]^{-1}, \quad \ldots\ldots \quad (4.9)$$

$$S_j = -\kappa \cdot \int_{S'} \hat{n} dS_q > 0. \quad \ldots\ldots\ldots \quad (4.10)$$

These Debye type temperatures Φ_j denote an average over a surface containing electrons with the *minimum* wave vector that can enter UK processes, in contrast to the usual Debye temperatures in which the maximum wave vector is sought. The 'effective surface' integral S_j is to be lumped with α_j and treated as an adjustable parameter.

No matter how we evaluate $I_s(T)$, eqn. (4.6) can be solved for I in terms of I_s:

$$I_j = T\left[\int^T \frac{1}{y} I_s(y) dy + C\right]. \quad \ldots\ldots\ldots \quad (4.11)$$

The term in the constant C is the solution of the homogeneous part of (4.6) C must be zero, as can be seen by evaluating (4.4) in the limit of extremely low T:

$$\lim_{T\to 0} I_j = -T\mathbf{K} \cdot \int dS_q \hat{n} \exp[-\hbar\omega/\kappa T]. \quad \ldots \quad (4.12)$$

But this corresponds to the first term of the bracket of (4.11). For it is decreasing exponentially, and there is no other way to make (4.11) equal (4.12) than by assigning $C = 0$. We then get with (4.8):

$$\frac{dI_j}{dT} = I_s + \int^I \frac{1}{y} I_s(y) dy$$

$$= S_j \left[\frac{z}{e^z - 1} + \int_z \frac{dx}{e^x - 1}\right] \ldots z = \Phi_j/T. \quad \ldots \quad (4.13)$$

Let us now write the integral S_j of (4.10) as a fraction γ_j times the integral taken over the surface S' of fig. 1:

$$S_j = \gamma_j S_{\max}, \quad \ldots\ldots\ldots \quad (4.14)$$

$$S_{\max} = -\int_{S'} K \cdot \hat{n}\, ds_q = 2{\cdot}824\pi k_0^3. \quad \ldots \quad (4.15)$$

Then if we define the function

$$F(z) = \frac{\Delta}{N}\frac{\kappa}{e}\frac{1}{2\pi^3} S_{\max}\left[\frac{z}{e^z - 1} + \int_z \frac{dx}{e^x - 1}\right]$$

$$= 367{\cdot}3 \left[\frac{z}{e^z - 1} + \begin{cases} -\ln z + \frac{z}{2} - \frac{B_1}{2!}\frac{z^2}{2} + \ldots z < 2\pi \\ \sum_{m=1}^{\infty} m^{-1} \exp(-mz) \ldots z > 1 \end{cases}\right] \frac{\mu V}{^\circ\mathrm{K}}, \quad (4.16)$$

we shall have from (4.2), (4.13), (4.14), and (4.16):

$$S_g{}^+ = \sum \alpha_j \gamma_j F\left(\frac{\Phi_j}{T}\right). \quad \ldots\ldots\ldots \quad (4.17)$$

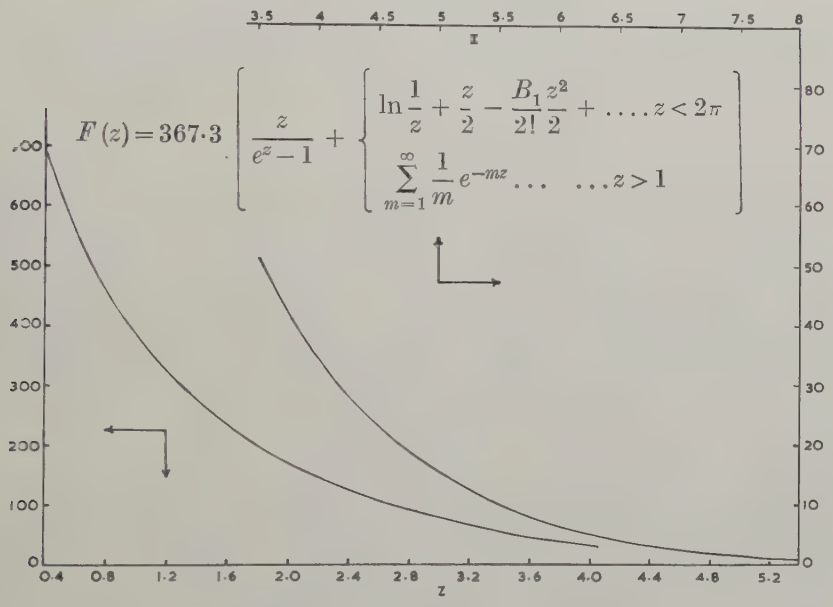
We have the result in the form of an 'effective α_j' times an 'effective surface ratio γ_j', times a function F of temperature. The combination

$$A_j \equiv \alpha_j \gamma_j \qquad \ldots\ldots\ldots \quad (4.18)$$

is treated as a constant adjusted to fit experiment, and the function $F(z)$ is shown in fig. 2.

This completes the derivation. It is not meant to be anything but a crude attempt at getting an expression which should represent in a fair way the temperature dependence of the phonon drag term S_g^+. The approximations (4.2), and (4.9) that went into this derivation were both of the same type, namely, it was assumed that quantities could be removed from under the integral sign at 'effective' values.

Fig. 2



The function $F(z)$ of eqn. (4.16).

To improve on the approximations we have made would require a detailed spectrum calculation for each metal at each temperature and a detailed knowledge of α as a function of q and K, a project beyond the scope of this paper. However, in the next section we begin a discussion of a less ambitious but similarly directed programme which is a 'baby spectrum' calculation incorporating only a small number of points on S′ and evaluated for only two sets of temperature and elastic constant parameters. From this we hope to get a glimpse of how qualitatively the contributions from the various polarizations j and the various modes **q** compare.

§ 5. The Calculation of Table 1

We here begin an analysis of the contributions to $S_g{}^+$. The analysis is made by selecting a number of q's on the surface S′ of fig. 1 and computing numerically for the three polarizations the factors

$$R_j = 10T\frac{\partial N(j)}{\partial T}\left(\frac{q}{k_0}\right)^2 \quad \text{and} \quad \alpha_j \quad . \quad . \quad . \quad . \quad . \quad (5.1)$$

and their product. The product here may be related either to the integrand of (4.1) or to the integrand of (4.11) (or of (4.13)). In the former case, we would need $\nabla_q\omega$ to complete the integrand, but we shall suppose that this linear factor will not upset any qualitative conclusions that are drawn from the exponential dN/dT and the α's. In the latter case, we notice that for sufficiently large z (i.e. low T),

$$R_j = 10\frac{z(q/k_0)^2}{(e^z-1)(1-e^{-z})} \cong 10\left(\frac{q}{k_0}\right)^2 zN_0. \quad . \quad . \quad . \quad (5.2)$$

This is precisely the integrand in (4.11) (or in (4.13)), see (4.7).

The points chosen were on the intersection of S′ and the six planes

$$\begin{aligned} &q_x = 0, \quad q_y = q_z, \\ &q_z = 0, \quad q_x = q_z, \\ &q_x = q_y, \quad q_x - q_y + q_z = 0. \quad . \quad . \quad . \quad . \quad . \quad (5.3) \end{aligned}$$

The results of the calculation are given in table 1. In addition to the qualitative conclusions about the relative importance of the various phonons, the numbers in table 1 enable us to make bounds for reasonable average values to be used for the Φ_j's. In this section we discuss the calculation that leads to table 1, and in the next section we discuss the conclusions and estimate the Φ_j. The first problem is the calculation of α.

From (2.2) we notice that the number of terms in the denominator of α will change from phonon to phonon depending on the number of UK processes in which a phonon can participate. If the number of types of interactions with electrons that a phonon can enter into is n, where $n = 1+m$, m being the number of UK processes, and the 1 referring to the normal process, then we call this an 'n-process phonon'. The very smallest q's then are '1-process phonons'. In table 1 the various entries are labelled by this parameter n. We originally thought to make a calculation for each 'n-process region' in the q-sphere, each region containing only 'n-process phonons'. But the actual differences in values for the α's when they were calculated were not that great to warrant such a computation.

If the numerator and denominator of (2.2) are divided by the numerator, we get

$$\alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K}_1) = \left[1 + \sum_{\mathbf{K}_2\neq\mathbf{K}_1}^{[\mathbf{q}j]} \beta(\mathbf{q}\mathbf{K}_1;\ \mathbf{K}_2)\right]^{-1} \ldots \frac{B_0}{B} \ll 1, \quad (5.4)$$

where, using eqn. (65) of Bailyn (1958) for the B's,

$$\beta(\mathbf{q}\mathbf{K}_1;\ \mathbf{K}_2) = \left\{\frac{\cos[\hat{\xi}(j\mathbf{q}),\ \mathbf{q}+\mathbf{K}_2]}{\cos[\hat{\xi}(j\mathbf{q}),\ \mathbf{q}+\mathbf{K}_1]}\right\}^2 \left|\frac{\mathbf{q}+\mathbf{K}_1}{\mathbf{q}+\mathbf{K}_2}\right| \left(\frac{I(\mathbf{q}+\mathbf{K}_2)}{I(\mathbf{q}+\mathbf{K}_1)}\right)^2. \quad (5.5)$$

Here I is the interaction matrix element of the electrons (eqn. (5) of Bailyn (1958)):

$$I(\mathbf{q}+\mathbf{K})=N\int_{\text{crystal}}\psi_{\mathbf{k}'}{}^{*}\nabla v\psi_{\mathbf{k}}d^{3}r,$$
$$\mathbf{k}'-\mathbf{k}=q+K, \qquad \ldots\ldots\ldots \quad (5.6)$$

where $\psi_{\mathbf{k}}$ is the wave function for the $\mathbf{k}$ electron, and $v(r)$ is the potential from an ion at $r=0$ plus the shielding electron cloud around it. $\hat{\xi}(j\mathbf{q})$ is the unit vector in the direction of vibration of the $\mathbf{q}j$ phonon. Since we are concerned mainly with the surface S′ of fig. 1, the I ratios in (5.5) will usually be either (*a*) between two UK processes with $|\mathbf{q}+\mathbf{K}_1|=2k_0$ and $|\mathbf{q}+\mathbf{K}_2|\cong 2k_0$, or (*b*) between one UK process and one normal process with $|\mathbf{q}+\mathbf{K}_1|=2k_0$ and $|\mathbf{q}+\mathbf{K}_2|=q=0{\cdot}2k_0\ldots.0{\cdot}5k_0$. In case (*a*) the I^2 ratio is about 1, and in case (*b*) the ratio is between about 1·0 (for Na) and about 0·5 (for Cs). The latter numbers are taken from the unpublished work of Bailyn (1956), but the effect on our results of using the Bardeen (1937) values for them would be approximately the same†. None of the qualitative results that arise from table 1 are sensitive to particular values chosen for parameters. Recent corrections taking into account exchange and correlation effects (to be published†) also do not alter the order of magnitudes of these ratios, nor consequently the final conclusions. The longitudinal mode contribution was the most sensitive to the I ratios, and that mode for more than one reason has such a small effect that in the end we neglected it altogether (see § 6 below).

The velocities $v(j\mathbf{q})$ and directions of vibration $\hat{\xi}(j\mathbf{q})$ were calculated using the elastic limit spectrum in the Appendix. The $\hat{\xi}$ in that spectrum are functions only of the 'isotropy parameter' s of the elastic constants c_{ij}:

$$s=\frac{c_{11}-c_{44}}{c_{12}+c_{44}}, \qquad \ldots\ldots\ldots \quad (5.7)$$

whereas the velocities depend on some additional elastic constant parameters. The elastic constants we used were calculated by the Fuchs method. The numbers cited by Huntington (1958) have since been improved using low-temperature compressibility data from the experiments of Swenson. (More details and data are discussed in a paper to be published†.) At the lowest temperatures (say 4°K), the values of s_0 are about 0·05, 0·15, 0·20, 0·23 and 0·27 for Li, Na, K, Rb, and Cs, respectively. s_0 is a small difference between large quantities, and very sensitive to the compressibilities, so these numbers are to be treated with caution. (By contrast $s_0=0{\cdot}7$ for gold, see Neighbors and Alers 1958). We made two complete calculations, one with $s_0=0{\cdot}1$ and the I^2 ratio in (5.5) equal to 1·0, the other with $s_0=0{\cdot}3$ and I^2 ratio equal to 0·5, only the latter given in table 1. These numbers were chosen as likely extreme situations. The $s_0=0{\cdot}3$ calculation would correspond roughly to Cs, whereas the $s_0=0{\cdot}1$ calculation we regard as very loosely corresponding to the sodium end

† *Added in proof.*—This work has since been published, see Bailyn (1960).

Table 1. First Half

Table 1. The 'baby spectrum' calculation for $s=0{\cdot}3$. n is the number of processes the phonon can engage in with electrons. $D_j=(\rho V_j^2/c_{44})^{1/2}$. The α's are the relative probabilities introduced in § 2. The α''s are the 'absolute probabilities' discussed in § 8. R is given by (5.1)

$q_z=0$	q_x	n	q/k_0	D_{I}	D_{II}	D_{L}	α_{I}	$\alpha_{\mathrm{I}'}$	α_{II}	$\alpha_{\mathrm{II}'}$	α_{L}	$\alpha_{\mathrm{L}'}$	$(\alpha R)_{\mathrm{I}}$	$(\alpha R)_{\mathrm{II}}$	$(\alpha R)_{\mathrm{L}}$
	0·0	5	0·30	1·00	1·00	1·33	0·50	(0·60)	0·00	(0·00)	0·10	(0·40)	0·22	0·00	0·02
	0·02	3	0·31	1·00	0·88	1·41	1·00	(0·60)	0·20	(0·10)	0·07	(0·30)	0·44	0·11	0·01
	0·4	3	0 34	1·00	0·64	1·54	1·00	(0·60)	0·48	(0·19)	0·05	(0·21)	0·44	0·47	0·01
	0·6	3	0·41	1·00	0·39	1·62	1·00	(0·60)	0·90	(0·26)	0·04	(0·14)	0·37	1·84	0·00
	0·71	5	0·49	1·00	0·32	1·63	0·25	(0·60)	0·25	(0·30)	0·03	(0·09)	0·09	0·79	0·00
	0·8	3	0·77	1·00	0·39	1·62	1·00	(0·60)	0·97	(0·37)	0·01	(0·03)	0·01	0·01	0·00
$q_x-q_yq_z$	q_x														
$=0$	0·0	2	0·20	1·00	0·32	1·63	0·00	(0·00)	0·00	(0·00)	0·17	(1·00)	0·00	0·00	0·03
	0·1	2	0·20	0·98	0·35	1·64	0·73	(0·02)	0·08	(0·00)	0·16	(0·98)	0·28	0·11	0·03
	0·2	2	0·21	0·94	0·43	1·64	0·78	(0·09)	0·03	(0·00)	0·16	(0·90)	0·33	0·03	0·03
	0·4	2	0·23	0·84	0·57	1·65	0·88	(0·21)	0·11	(0·02)	0·16	(0·77)	0·47	0·10	0·02
	0·6	3	0·31	0·95	0·41	1·64	0·76	(0·41)	0·52	(0·29)	0·07	(0·30)	0·36	0·87	0·01
	0·71	5	0·49	1·00	0·32	1·63	0·25	(0·60)	0·25	(0·30)	0·03	(0·09)	0·09	0·79	0·00
$q_x=q_y$	q_z														
	0·0	5	0·49	1·00	0·32	1·63	0·25	(0·60)	0·25	(0·30)	0·03	(0·09)	0·09	0·79	0·00
	0·1	4	0·40	0·99	0·34	1·64	0·45	(0·64)	0·50	(0·30)	0·02	(0·06)	0·19	1·29	0·01
	0·2	4	0·34	0·94	0·36	1·66	0·33	(0·67)	0·50	(0·30)	0·01	(0·03)	0·16	1·05	0·01
	0·4	4	0·28	0·79	0·50	1·70	0·19	(0·70)	0·50	(0·30)	0·00	(0·00)	0·15	0·71	0·00
	0·58	4	0·25	0·63	0·63	1·72	0·67	(0·40)	0·00	(0·00)	0·10	(0·60)	0·55	0·00	0·01
	0·6	3	0·25	0·62	0·65	1·72	0·13	(0·68)	0·50	(0·30)	0·00	(0·02)	0·11	0·39	0·00

Table 1. Second Half

$q_x=0$	q_z	n	q/k_0	D_{I}	D_{II}	D_{L}	α_{I}	$\alpha_{\mathrm{I}'}$	α_{II}	$\alpha_{\mathrm{II}'}$	α_{L}	$\alpha_{\mathrm{L}'}$	$(\alpha R)_{\mathrm{I}}$	$(\alpha R)_{\mathrm{II}}$	$(\alpha R)_{\mathrm{L}}$
	0·0	5													
	0·2	2	0·24	1·00	0·88	1·41	0·00	(0·00)	0·23	(0·09)	0·11	(0·91)	0·00	0·07	0·02
	0·4	2	0·22	1·00	0·64	1·54	0·00	(0·00)	0·03	(0·02)	0·10	(0·98)	0·00	0·02	0·02
	0·6	2	0·21	1·00	0·39	1·62	0·00	(0·00)	0·02	(0·00)	0·11	(1·00)	0·00	0·02	0·02
	0·71	2	0·20	1·00	0·32	1·63	0·00	(0·00)	0·00	(0·00)	0·17	(1·00)	0·00	0·00	0·03
$q_y=q_z$	q_x														
	0·0	2													
	0·1	2	0·20	0·98	0·34	1·64	0·63	(0·03)	0·00	(0·00)	0·16	(1·00)	0·30	0·00	0·03
	0·2	2	0·20	0·94	0·37	1·66	0·72	(0·09)	0·00	(0·00)	0·16	(0·99)	0·30	0·00	0·03
	0·35	2	0·21	0·83	0·46	1·69	0·85	(0·21)	0·00	(0·00)	0·18	(0·99)	0·43	0·00	0·03
	0·50	2	0·23	0·70	0·58	1·72	0·90	(0·40)	0·00	(0·00)	0·19	(0·99)	0·62	0·00	0·03
	0·58	4													
$q_x=q_z$	q_y														
	0·58	4													
	0·6	3	0·25	0·63	0·63	1·72	0·50	(0·30)	0·02	(0·03)	0·18	(0·67)	0·41	0·02	0·02
	0·7	3	0·24	0·75	0·54	1·71	0·50	(0·30)	0·50	(0·11)	0·11	(0·58)	0·32	0·48	0·02
	0·85	3	0·23	0·87	0·60	1·63	0·50	(0·30)	0·17	(0·02)	0·12	(0·68)	0·26	0·14	0·02
	0·95	3	0·25	0·95	0·79	1·49	0·50	(0·30)	0·00	(0·00)	0·14	(0·70)	0·23	0·00	0·03
	1·00	5													

of the alkalis, even though the s_0 of sodium is considerably higher. The point is that qualitatively these two calculations give the same result, and we concluded that this result will be valid for the actual metals. There is this quantitative difference, however, that the α's which are relatively large in the $s_0 = 0{\cdot}1$ calculation become relatively smaller in the $s_0 = 0{\cdot}3$ calculation. A rise in the matrix element I^2 ratio makes the α's smaller.

Finally we wish to point out an important distinction between the directions 011 and 110 in table 1. When (4.1) is reached, the sum over **K** in (3.3*b*) has been reduced to just a term in one **K**. This **K** we chose for the purpose of computing table 1 to be the vector whose tail is at the reciprocal lattice point 011 and whose tip is at 000. The **q**-vector then starts at 000 and proceeds to some point on S′. The surface S′ is thus centred about $-\mathbf{K}$; hence we shall refer to it from now on as S′ (011). Now once this direction is chosen for **K**, one should be careful to note that the 011 direction is the one and only direction of a phonon q on S′ (011) that corresponds to an interaction in which an electron **k** flips from a direction (011) of closest approach of the Fermi surface to the Brillouin Zone boundary to the opposite direction (011). There are other directions in a cubic crystal which are equivalent to the 011 direction (such as for example the 110 direction), but *relative to the surface* S′(011), these directions are obviously no longer equivalent. We mention this here since we shall find that on S′(011), the 110 direction is quite important, whereas the 011 direction is not. But the 110 phonons here do *not* correspond to interactions in which an electron in the direction of closest approach to the Brillouin Zone boundary flips to the opposite direction.

§ 6. Discussion of Table 1

Our first conclusion from table 1 is that the longitudinal mode contributes far less than either of the transverse modes, *both* because the longitudinal α's are small and because the longitudinal velocities v_{L} are large. We shall in fact neglect the longitudinal mode. It is interesting to note that the closer q gets to the 011 direction, the larger the contribution gets in this mode. The 011 direction corresponds as mentioned above to electrons in the direction of closest approach to the Brillouin Zone flipping to the opposite direction. If the longitudinal were the only mode contributing, the distance of closest approach would thus be as crucial a matter as hypothesized by Ziman (1959). S_g is the very small difference between the large numbers $S_g{}^-$ and $S_g{}^+$, and it is possible that the small longitudinal contribution could still be significant. But it is so much smaller than the others, even for its relatively important directions that we can not see how we can pretend to an accuracy that would include it.

Our second conclusion is that in the transverse modes, the most important phonons are not the ones with the smallest wave vector q (i.e. not the 011 phonons). We can therefore not rely on the change in q magnitude to give the correct relative importance of the different *directions* in the shaded volume in fig. 1. (But we should expect to be able to rely on the change in

q to give the correct relative importance of phonons in a given direction, as q magnitude gets larger in that direction.) The transverse I mode (the one with the larger velocity) has appreciable contributions apparently from a considerable part of S′. We find that in a very rough way, an upper limit on q/k_0 (on S′) is about 0·50 which makes the γ_{I} of (4.14) approximately 0·15.

On the other hand, a much more restricted region of S′ seems to be important for the transverse II mode. This region contains the 110 direction significantly, and we estimate in a very rough way that it includes q/k_0's from about 0·35 to about 0·50. Such a region provides $\gamma_{\mathrm{II}}=0{\cdot}09$ approximately. However, the very erratic nature of the contributions from this mode seem to indicate that the contributions may not be sustained as the azimuthal angle covers its radians (about the polar axis which is **K**). It would not be unreasonable to expect that possibly only a fraction of $\gamma_{\mathrm{II}}=0{\cdot}09$ would be in fact contributing. Our attempts at fitting experiment place the adjustable parameter $\gamma_j\alpha_j$ at round 0·03 for this mode (varying slightly from metal to metal). The smallness of this number could be interpreted either as a result of small γ_{II}'s or small α_{II}'s. It appears to us much more consistent with table 1 to believe that the smallness arises from the small participating region rather than from the smallness of α in that region.

The transverse II mode contains phonons (near the 110 direction) which *individually* contribute to $S_g{}^+$ more than any other phonons from any mode. It strikes us as quite surprising that the q's of these phonons turn out to be as large as $0{\cdot}49k_0$, that is, well above twice the minimum wave vector magnitude ($0{\cdot}2k_0$) that a phonon participating in an UK process can have.

The preceding discussion and computation was based on a spherical Fermi surface. If we contemplate a bulging of the Fermi surface toward the Brillouin Zone boundary at the direction of closest approach, then in fig. 1, this would correspond to the shaded volume bulging toward $q=0$. Such a bulging is most important for the 011 phonons (on S′ (011)), i.e. for the directions in which q is the smallest. Accompanying every bulge, there will also be an indentation somewhere on the Fermi surface, since the volume in k-space occupied by the electron distribution is the same no matter what the details of shape are. Eventually therefore, there will be a 'negative bulge effect' for electrons which flip in an UK process from one 'negative bulge' to the opposite direction. Such processes will usually require rather large q's. For the phonons which are important in the transverse modes and especially in the transverse II mode, we estimate that the 'positive' bulge effect is diminished from what one would expect if the most important phonons were in the 011 direction. We do not however anticipate that the 'negative-bulge' effect would have arrived for these phonons, as their q's are still small despite the fact that they are of the order of twice the minimum q's. In our crude result (4.13) the effect of a bulge could be taken into account by altering the values for the

effective Φ_j's. Thus the Fermi surface distortion effect will be indistinguishable from just a different Debye temperature Φ_j† or a difference in spectrum characteristics. This seems to us a rather general result, and we regard as premature the hope of using phonon drag effects to indicate the degree of distortion of the Fermi surface. In our computations, as we went from Na to Cs, we have not postulated any unusual decreases in the Φ_j that would correspond to the progressive increase in distortion of the Fermi surface. (Small errors in Φ arising from say an inaccurate c_{44} or spectrum average errors or Fermi surface distortion can be indirectly and partially corrected by adjusting $\alpha\gamma$ (see below).)

§ 7. Numerical Computation

We have emphasized above that the calculations here are very crude, and the computed quantity very sensitive to parameter values; and we regard our attempt at fitting experiment as showing little more than that to within the crudeness of the computations, it is possible to get qualitatively correct curves over some of the temperature range. Ziman (1959) was the first to show that, despite the difficulties, numerical computations of at least some significance could be made.

The parameters Φ_j are decided upon in a 'consistent way' for the four metals Na, K, Rb, and Cs. They are defined symbolically from the Appendix by

$$\langle z_j \rangle \equiv \frac{\Phi_j}{T} = \frac{\theta}{T}\left\langle \frac{q}{k_0}\left(\frac{\rho V_j^2}{c_{44}}\right)^{1/2}\right\rangle_{\text{average}},$$

$$\theta = \frac{\hbar k_0}{\kappa}\left(\frac{c_{44}}{\rho}\right)^{1/2}, \qquad (7.1)$$

the symbolic aspect being the nature of the average (i.e. it is a *function* of z that is actually averaged in (4.9)). ρ here is the density and κ the Boltzmann's constant. From our two calculations at $s=0{\cdot}1$ and $s=0{\cdot}3$ it appears that the product

$$\frac{q}{k_0}\left(\frac{\rho V_j^2}{c_{44}}\right)^{1/2} = \frac{\Phi_j(\mathbf{q})}{\theta}$$

does not vary much from metal to metal (i.e. as s_0 varies) for each direction $\hat{\mathbf{q}}$. Thus our 'consistent way' mentioned above is to allow, for each polarization, Φ_j/θ to take on for each j the same average value for all the metals. We have found that $\Phi_{\mathrm{II}}/\theta=0{\cdot}25$ and $\Phi_{\mathrm{I}}/\theta=0{\cdot}35$ enables us to get fair results. This would correspond for example to an average q/k_0 of about 0·4 for both polarizations and an average $(\rho V_j^2/c_{44})^{1/2}$ of about 0·90 for the transverse I vibrations and of about 0·62 for the transverse II vibrations. These numbers do not seem unreasonable in view of table 1. The θ's are calculated using computed elastic constants (see table II and Huntington 1958‡). In metals with higher θ the average in (7.1) at a given T will emphasize smaller $\Phi(\mathbf{q})$'s. Thus we expect that use of the same

† This point was also noted by Ziman (private communication).
‡ See also Bailyn (1960).

average Φ_j in all the metals (for each j) would in fact be giving a value which is too low for Cs as compared with Na. Such an effect would tend to offset the Fermi surface distortion effect. In any case it would seem to be small, and we have not attempted to take it into account.

The results of the computations are given in fig. 3, and some of the data in table 2. The first problem in comparing theory with experiment is to estimate the specific heat, which is required for S_g^-. Recently there have been measurements of this quantity for Li Na and K by Roberts (1957) in the temperature range 2–20°K. We notice that the Debye temperature computed by Roberts for K goes through a sizeable minimum at about $T = T_{\min} \cong 6$°K. We interpret this result as an indication of the fact that as T decreases, the transverse II mode, with its lower velocities, tends to become more important. As T drops still further, the Debye temperature begins to rise. This is usually explained by supposing that for very small q, the density of states varies more swiftly than q^2 (which the Debye theory gives). Our crude reduction of a spectrum to two Φ_j's

Table 2. Data used in the computations. The first seven rows give the values of the Debye temperatures of the specific heat

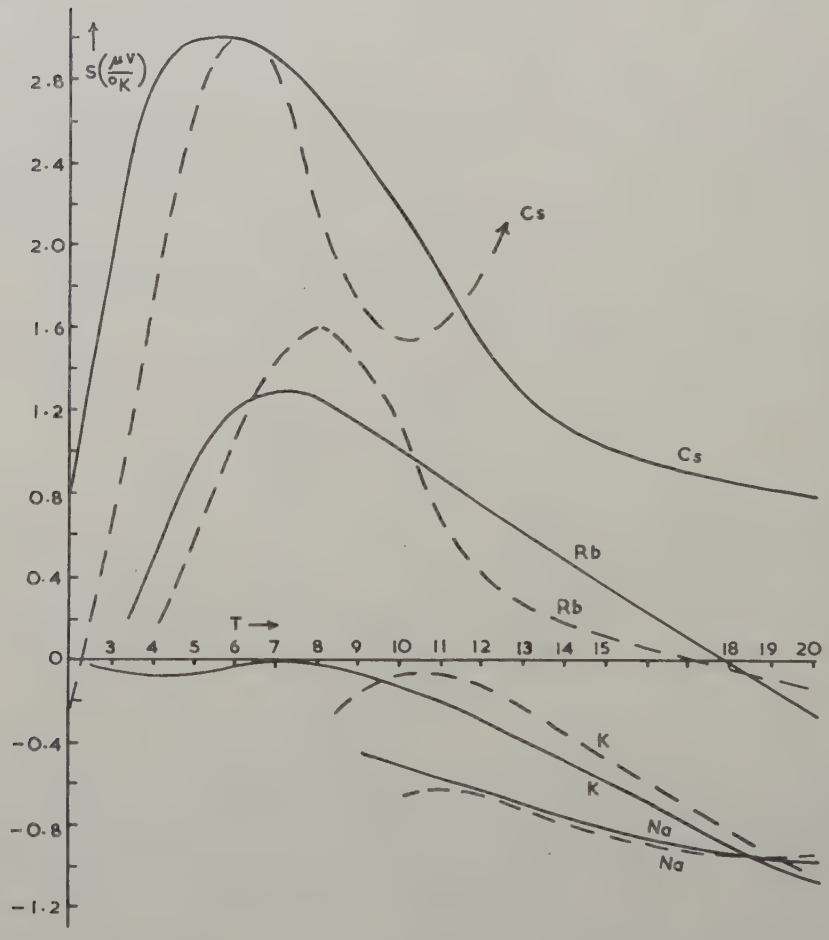
	Na	K	Rb	Cs
$\Theta(4°K)$		84·4	54	41·5
$\Theta(6°K)$		82·6	54	42·0
$\Theta(8°K)$		84·5	55	42·5
$\Theta(10°K)$	(155)	87·5	56	43·0
$\Theta(12°K)$	150	89·5	57	43·5
$\Theta(16°K)$	156	94·0	58	44·0
$\Theta(20°K)$	163	96·0	58	44·0
$\alpha_I\gamma_I$	0·07	0·11	0·12	0·12
$\alpha_{II}\gamma_{II}$	0·035	0·029	0·028	0·030
Φ_I	60	34	20	15
Φ_{II}	43	24	15	11
θ K	172	96	58	44

would enable us to provide the first effect, namely the lowering of the Debye temperature in the specific heat as T decreases, but it would be incapable of providing the second effect, the rise in the Debye temperature as T goes still lower. Thus the temperature range in which with our approximations to the spectrum we have at least a hope of getting a fit with experiment is the range above the minimum position. We shall attempt to fit, therefore, only this range for each metal. In the case of sodium, on the other hand, the specific heat experiments give a fairly constant Debye temperature. The explanation for this does not seem to have been determined†. Various theoretical curves for Bauer and Bhatia and Horton have been quoted in a review article by Parkinson (1958),

† See however footnote 24 of Bailyn (1958).

and all of them give a minimum in the Debye temperature more or less like what one finds experimentally for potassium, but not for sodium. Since the theory is unable to predict the specific heat of sodium, it seems unreasonable to suppose that it can predict the even more delicate quantity, the thermoelectric power. We can try however to use a consistently theoretical calculation, even though the specific heat part does not yield the observed specific heat behaviour. Thus we are led to use

Fig. 3



The thermoelectric power of the alkali metals. The dashed lines are the theoretical results, the full lines are the experimental. The wild high temperature caesium result is interpreted as caused by neglect of the phonon-phonon interactions ; and the low temperature limitations on Na and K result from use of a spectrum which does not have a density of state function more accurate than the Debye approximation $\sim q^2$ for small q. The discussion is in § 7.

one of the theoretical curves cited by Parkinson for computing S_g^-. We have in fact used Bauer's curve D shown in fig. 12 of Parkinson's article. The curve has a minimum at about 11°K, so we have calculated results only for T greater than this. We can get some agreement as shown in fig. 3, but we do not dare to call it significant. The behaviour of sodium is surely not 'ideal'. This might be due in part to the martensitic transformation it undergoes at low temperatures.

For rubidium and caesium, there do not seem to be specific heat measurements over the necessary temperature range at all, so we manufactured specific heats by analogy with the observed potassium one. We anticipate that the specific heat Debye temperature will show a minimum in Rb and Cs just as in K, and that it will occur at a temperature which is proportional to the corresponding temperature in K, the constant of proportionality being θ_{Cs}/θ_K. We also shall regard the size of the minimum as being proportional to the size in K, with the same constant of proportionality. In this way we have constructed the specific heats for Rb and Cs that are given in table 2. We do not believe that more elaborate theoretical computations would be especially significant without more experimental data. It should be emphasized that we should also expect a similar behaviour in sodium, and in fact our specific heat taken from Bauer's work does show something of this sort of behaviour. (Bauer's minimum occurs at $T \cong 11$°K and $(\theta_{Na}/\theta_K)T_{\min K} \cong 11$°K). If we are consistent we must regard sodium as behaving anomalously, and that in fact is our view, as mentioned above.

The results of our calculation of S_g for the other metals are as follows. In potassium, the observed thermoelectric power hovers about an extremely small value as T approaches zero. We are able to match the observed values down to about 9°K. Below this our curve becomes considerably too large negatively. We attribute this, to the failure of our spectrum below the Debye Θ minimum in the specific heat. For potassium this should start in at around 7°K, (the minimum is reached at ~ 6°K). The remaining disagreement reflects the poorness of our approximations and perhaps the phonon–phonon interactions which may, even though small, have an important effect, since the resultant S_g is such a small difference between large numbers.

The rubidium and caesium curves are fairly successful. In the latter, however, above about 11°K, the calculated curve gets out of control. We attribute this to the decisive onset of phonon–phonon interactions. When these become large, the α's become temperature dependent, and another calculation entirely must be contemplated. The first place we would expect this drastic change in S_g would be in the higher temperatures for caesium, since caesium has the lowest θ. We therefore tentatively ascribe to this effect the wildness in the calculated curve for caesium above about 11°K.

It was partially the fear that this type of difficulty would interfere with all the calculations that caused us to be skeptical at first that any significant calculations could be made. Ziman (1959) was the first to show that

calculations could be made by neglecting the phonon–phonon processes entirely. However, we suggest that they are entering finally in a decisive fashion here in the higher temperature range for caesium.

The small changes in the $\alpha\gamma$ parameter in going from K to Cs are quite essential in getting a fit with experiment. Those changes reflect the changes in going from metal to metal, and also indirectly the changes in Φ that are not taken into account by our 'systematic' method of determining Φ, such as possible erroneous c_{44} values, distortion of the Fermi surface, and spectrum details, all of which influence the same parameters Φ. There is experimentally an increasingly positive character of S_g as we go from Na to Cs that we have tried to match. Unfortunately we cannot trace the individual effects, and show how they would produce this result: all that we are able to do is provide a form for the result in terms of adjustable parameters. Qualitatively we can suggest the following possible origins of the increase in positive character in going from Na to Cs.

(A) The spectrum shape (the $\Phi(\mathbf{q})$'s) changes in a way that indicates increasing $S_g{}^+$ as s_0 increases; and we have found a general tendency for s_0 to increase as we go down the alkalis. What is important is the possibly different effect on $S_g{}^-$ from on $S_g{}^+$ caused by a change in the spectrum shape. Our results indicate that the temperature variation in the average Debye temperature of the specific heat ($\Theta(T)$) gives a large effect, and a difference in the behaviour of $\Theta(T)$ from the behaviour of an effective average $\Phi(T)$ (representing our two Φ's) owing to a change in the spectrum shape might well yield significant changes in the resultant S_g. It is clear that to pursue this point would require enormous calculations (note that simply a change in θ (eqn. (7.1)) would have a similar effect on $S_g{}^-$ as on $S_g{}^+$).

(B1) Under this heading and the next we shall be appealing to calculations which at the time of writing are unpublished but which should appear in the near future†. These calculations consist in an improvement of the matrix element, and a detailed calculation of the electrical and thermal conductivities using a phonon spectrum on the Born–von Karman model containing up to next-nearest neighbour interactions. The results of the matrix element calculation indicate that the I^2 ratio (in the discussion after eqn. (5.6)) decreases considerably as we go from Na to Cs. This results not only from the shielded ion perturbation but also and quite significantly from the exchange and correlation correction, which we have calculated numerically for the first time. Unfortunately the latter was done only for the three cases $r_s/a_h = 3$, 4 and 5, but we find that the increase over the 1937 Bardeen value of the I^2 ratio is 1·54, 0·91 and 0·65 for these three cases respectively. We assume that this tendency will continue for $r_s/a_h = 6$. What this means is that there is a significant relative increase of the (UK/non-UK) ratio as we go from Na to Cs, an increase that is completely absent from the original Bardeen calculation.

† *Added in proof.*—Now published: Bailyn (1960).

In addition there is a similar increase for the shielded-ion part itself. This could have a considerable effect, but as above, the calculation that would trace it through would have to be tremendous.

(B2) The results of our calculations of the electrical and thermal conductivities indicate that the highly anisotropic spectrum that we have calculated is fairly good for caesium but gets increasingly worse as we go from Cs to Na. This conclusion is based on the fact that our results as the temperature decreases are fairly good for Cs but get progressively worse for Rb, K and Na, the poorness being a too-large value of the calculated resistivity. These large values are the consequence of the few very small velocity phonons. The conclusion then is that these velocities in the case of sodium are in fact not so small. If in fact the high anisotropy as we have calculated it should be damped down as we go from Cs to Na, this means that the relative importance of the UK to non-UK effects should get similarly progressively damped and in a way not contemplated by the Born–von Karman spectrum including interactions up to next-nearest neighbours. Our spectrum analysis as done above would then be justified for caesium, but less so for sodium. This would also mean that considerable work lies ahead in getting a better spectrum for sodium. Our result here ties in with our considerations on the specific heat in sodium. Our feeling is that this aspect of the spectrum problem is the main source of the really great increase in the positive character of the caesium result (relative to rubidium, potassium and sodium).

(C) The progressive distortion of the Fermi surface would also seem to yield an increasingly positive S_g as we go from Na to Cs. In our rough expression (4.17) it would have the effect of lowering the Φ values, or in an indirect way raising the α's, since a lowering of Φ over a small temperature range can be imitated by raising of the $\alpha\gamma$ parameter. Ziman has interpreted the entire change in the positive character in going from Na to Cs as a consequence of this effect. Our point of view is that the many effects mentioned above very likely play even more important roles, and that it is hard if not impossible to distinguish one from another, since they may all be regarded as changing effective Φ values or effective $\alpha\gamma$ values. It is only by a very detailed calculation that one could eliminate the possibility of, or establish upper bounds to, the alterations caused by these effects. We therefore regard Ziman's conclusions with caution if only for his omitting such a detailed calculation. There are however more important criticisms we have of his method, which we shall discuss in the next section.

§ 8. Comparison with the Work of Ziman

Since Ziman (1959) has put forth a calculation some time ago concerning the same quantity as discussed in this paper, it seems appropriate to compare his procedure with ours. We mentioned in the Introduction that we had objections to his method, and in this section we shall outline what they are. They may be separated into the three statements: (i) his

variational principle is essentially weaker than ours (Bailyn (1958)†) and leads to what we consider an erroneous and rather complicated although perhaps approximate expression for the phonon–drag component to the thermoelectric power (i.e. eqn. (2) of Ziman (1959) compared to our eqn. (2.8)); (ii) his variational principle does not give any indication of the relationship between the hypotheses for the non-equilibrium part of the phonon distribution and the non-equilibrium part of the electron distribution, with the result that Ziman chose ansätze (eqn. (5) of Ziman 1959) which were inconsistent with the Boltzmann equations and which do not reflect the relative probabilities α which we have emphasized are characteristic of phonon-drag effects, and should appear in these distributions (see below); and (iii) in using a very simple spectrum for the phonons, he makes a computational error of qualitative significance, for he was led to suppose that the effect of the distortion of the Fermi surface was rather clear cut, and could be taken into account in a simple way, whereas in actual fact, the effect of the distortion of the Fermi surface should be considered in its effect on the *transverse* modes (the longitudinal one not being important) which are highly anisotropic (with the most important phonons not the ones with minimum q) and do not lend themselves to the simple calculation Ziman proposed. We shall comment further on the first two points.

(i) Ziman's variational principle was enunciated in 1956. His method was to add the 'entropy production' term from the phonon Boltzmann equation to that of the electron Boltzmann equation in a certain manner. When this was done he found that the resultant expression obeyed the conditions for the applicability of the Variational Principle with respect to variations either in the phonon distribution or the electron distribution. Our method was to assume the existence of a time of relaxation for the phonon–phonon collision term, solve the phonon Boltzmann equation for the non-equilibrium component of the phonon distribution in terms of the non-equilibrium electron distribution and then substitute that solution into the electron Boltzmann equation, which thereupon lent itself to a variation principle. In this way, there was only one unknown (the electron distribution). The convergence of the series (for the electrical resistivity) was the same as in the original work of Kohler. It is not clear to us what can be said about the convergence in Ziman's treatment. It is, nevertheless, curious that we have come out with expressions for S_g that are quite different. To be able to compare the methods in greater detail, we seek what connection our approach might have with an entropy of some sort. The connection can be obtained by following the interpretation of eqn. (14) of BI that is given at the end of § 3 of that paper. We rewrite this generalized Boltzmann equation symbolically as

$$\left(\frac{\partial f}{\partial t}\right)_{\text{field}} + \frac{\partial f}{\partial N}\left(\frac{\partial N}{\partial t}\right)_{\text{field}} = -\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} - \frac{\partial f}{\partial N}\left(\frac{\partial N}{\partial t}\right)_{\text{elec. coll}}. \qquad (8.1)$$

† This article will be referred to as BI from hereon.

Here N represents the phonon distribution. The first term on the left is the alteration of the electron distribution $f(\mathbf{k})$ resulting from the direct action of the fields on it. The second term on the left (which in BI is denoted by U) is the alteration resulting from the indirect action of the fields on the electrons, through the phonons. (i.e. The fields will produce a non-equilibrium component in the phonon distribution which in turn passes some of this on to the electrons. This is the customary notion of 'phonon drag'.) Thus the two terms on the left provide the *total effect of the external fields* on the electrons which we indicate symbolically as

$$\left(\frac{df}{dt}\right)_{\text{field}} = \left(\frac{\partial f}{\partial t}\right)_{\text{field}} + \frac{\partial f}{\partial N}\left(\frac{\partial N}{\partial t}\right)_{\text{field}} . \quad . \quad . \quad . \quad (8.2)$$

The first terms on the right of (8.1) (which in BI is denoted by L_1) gives the change in the electron distribution arising from the collisions with the equilibrium part of the phonon distribution. The second term on the right (which in BI is denoted by L_2) gives the change in the electron distribution arising from collisions with that part of the non-equilibrium component of the phonon distribution that was produced by interaction with the electrons. Thus the two terms of the right provide the total change in the electron distribution arising from collisions with the phonons, when the indirect effect of the fields is subtracted off, which we represent symbolically by

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \left(\frac{df}{dt}\right)_{\text{coll}} + \frac{\partial f}{\partial N}\left(\frac{\partial N}{\partial t}\right)_{\text{elec. coll}} . \quad . \quad . \quad . \quad . \quad (8.3)$$

Thus our generalized Boltzmann eqn. (8.1) for the electrons is of the same form as the ordinary equation, and one can go about constructing entropy production in the usual way, and end up with precisely our variational principle. Thus, whereas Ziman constructs what he calls the total entropy production of the system, we construct what is a kind of *partial* entropy production, referring to the electrons. An entropy production for the electrons alone could be constructed as a product of affinities and fluxes in the way Ziman (1956) proposed, but the temperature affinity $X_k(\text{temp})$ (analogous to the electric field affinity in eqn. (25) of Ziman's paper) would have to reflect the left-hand side of (8.1):

$$X_k(\text{temp}) = -\frac{1}{T}\frac{dT}{dx}\left[-\frac{\hbar}{m}(E-\rho)k_x\frac{\partial f_0}{\partial E} - \sum_{qK}\frac{\partial N_0}{\partial \hbar\omega}\hbar\omega(jq)\frac{\partial \omega(jq)}{\partial q_x}\,\alpha(jq;\ q+K)\right].$$

(To get this result, one needs to evaluate the U of eqn. (14) of BI.) To complete the analogy with the formulation of Ziman, one needs to show that the operator on the right-hand side of (8.1) satisfies the correct symmetry conditions. That, however, was the burden of BI and has been shown.

It should be emphasized that we regard as the real justification of our results the statistical mechanical argument of BI. The thermodynamic garb is produced here solely to compare with Ziman's work.

There is the question however: *ought* there to be maximizing of the entropy of the electron system alone? We believe the answer to this is yes, for the reason that the electron–phonon interaction still results by assumption in a first-order deviation of the electrons from equilibrium, and one may therefore legitimately lump the non-equilibrium phonons in with the reservoir, and apply fluctuation theory to the description of the electron system in the environment of this complex reservoir. It then becomes a problem what to use for affinities, however, and the usefulness of this as a general principle disappears. On the other hand, if one follows the Ziman-type approach, one can define entropy production as Ziman in (1956) using the above affinity. Just what the relation between the Ziman-type entropy and the usual notion of entropy is has not been made altogether clear (see the comment of Prigogine at the end of the paper by Ziman (1956)). Perhaps the safest way is to regard Ziman's as simply a maximizing function. What we have shown in BI is that a stronger maximizing function exists the solutions from which cast doubts on the significance of Ziman's expressions. Thus, we have a much more restricted class of variations possible than does Ziman. The exact solutions of the Boltzmann equations will maximize both entropy productions, but our view is that if what is desired is a method for obtaining approximate solutions to the Boltzmann equations, then our method with a more restricted range of variations will yield a solution which converges more swiftly than the method with the larger range of variations. We are not sure that there might not exist some other quantity with an even more restricted range which would yield a better convergence still, but the fact that in the present paper we were able to derive our expression from a rather simple 'mechanical' model leads us to believe that there is, in fact, not. We thus regard the Ziman expression as a term whose significance is not obvious and which at best would be equivalent to ours.

The original restriction on our method, namely, that phonon–phonon interactions could be written in terms of a relaxation time, has been removed by Tsuji (1959), who showed that a generalization can be made whenever a certain inverse collision operator exists.

(ii) To get the best results from Ziman's method, one would expect to consult the Boltzmann equations to obtain the trial phonon distribution once the electron trial distribution is decided on. In a variational technique, one can avoid doing this and hope for the best, but when a relationship does exist, one would expect the *best* results from using it. Ziman chose the electron trial non-equilibrium distribution as a constant times $\mathbf{k}$. It then follows immediately from the phonon Boltzmann equation (see eqns. (12), (13) and (69) of Bailyn 1958) that the corresponding trial phonon non-equilibrium distribution is obtained from $N = N_0 - (\partial N_0/\partial \hbar\omega)G$ where

$$G = \frac{\tau_c \hbar\omega \nabla_q \omega}{\sum_{\mathbf{K}} B(j\mathbf{q};\ \mathbf{q}+\mathbf{K})} \frac{1}{T}\frac{dT}{dx} - \text{const.}\left[\sum_{\mathbf{K}}^{[j\mathbf{q}]} (\mathbf{q}+\mathbf{K})\alpha(j\mathbf{q};\ \mathbf{q}+\mathbf{K})\right], \quad (8.4)$$

provided the phonon–phonon interactions are neglected. Here α is defined as in this paper and τ_c is a constant given by eqn. (70) of BI.

What Ziman uses for G instead is a constant times $\mathbf{q}$, which may be interpreted as an approximation to the first term of (8.4) (which arose from the drift term of eqn. (12) of BI). Such an approximation would not necessarily be a calamity, but in the phonon-drag effect, we have a resultant S_g which is a small difference between a large positive UK contribution and a large negative non-UK contribution. Now what Ziman neglects in (8.4) in particular are all the 'UK components' in the phonon distribution. We therefore do not see how a computation based on these trial distribution functions can be significant†.

It is a strange result however, that in our theory the terms in the square bracket of (8.4) affect only quantities (the d_{ij}'s of BI) which cancel out in our expression for S_g (see eqn. (86) of BI). Thus on our theory, the only part of (8.4) that finds itself in S_g is the term Ziman uses. Thus Ziman's guess for G is in a sense consistent with our theory but not with his own!

Finally, even if one accepts Ziman's expression, and accepts his neglect of the square bracket in (8.4), there is the denominator in (8.4) to be considered. In every interaction there will be an interaction matrix element which is proportional to

$$\alpha'_j(\mathbf{q})I(\mathbf{q}+\mathbf{K})^2/|\mathbf{q}+\mathbf{K}|, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.5)$$

where

$$\alpha'_j(q)=\left[\hat{\xi}_j(\mathbf{q})\,.\,\frac{\mathbf{q}+\mathbf{K}}{|\mathbf{q}+\mathbf{K}|}\right]^2. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (8.6)$$

Thus in the integrals, there will always be the first term of (8.4) times (8.5) from which it can be seen that a factor $\alpha(j\mathbf{q}\,;\ \mathbf{q}+\mathbf{K})$ will always emerge. If one neglects to consider this, then in effect one is replacing α by α'. (The factors I and $|\mathbf{q}+\mathbf{K}|$ in (8.5) will be constant over the important region, i.e. over the surface S′ of fig. 1, on which $|\mathbf{q}+\mathbf{K}|=2k_0$.) To see the consequence of this, we have placed beside each column of α's in table 1, a corresponding column of α''s. We notice that the α''s emphasize the small q's of the longitudinal mode quite considerably. Thus for still a third reason, we are doubtful of the significance of Ziman's treatment.

§ 9. Summary

In this paper we have applied the theory of Bailyn (1958) to the case of the thermoelectric power of the alkali metals in the temperature range 2–20°K. A simple derivation of the basic equation was shown to be possible. It turns out that quantities that we have called 'relative probabilities' enter importantly in the drag effects. In order to see their characteristics, they were computed for several points by means of a simple spectrum using values for parameters that should reflect at least roughly the situation in the alkali metals, and neglecting the effect of

† We are of course relying here on the equations of Ziman (1959) in which (the total) G appears (our G is proportional to Ziman's Φ_g). It is *possible* that the sum in (8.4) is negligible, but that would have to be shown.

phonon–phonon collisions. The conclusion drawn was that the longitudinal vibrations contributed so small an amount relative to the transverse vibrations that they were negligible. The remaining transverse vibrations betrayed the surprising characteristic of having the most important phonons not the ones with minimum q in the UK process, even in the low-temperature range considered. Treating the relative probability as a constant removable from under the integral sign (a *very* crude approximation), we were able to compute rather simple expressions for the phonon-drag of the thermoelectric power. From these expressions, numerical results were computed which indicate that the experiments are at least not inconsistent with the theory in its present state of crudeness. Our main attempt here has been to establish the basic equations, and to analyse qualitatively the nature of the contributions to the effect. At the end, we have stated our objections to a previous calculation of Ziman which also calculated the phonon-drag effect in the thermoelectric power of the alkali metals.

Acknowledgments

I wish to thank Dr. D. K. C. MacDonald for his stimulation and encouragement on this problem. Without it, this investigation would never have been undertaken or brought to the present point. I should also like to thank the Low-Temperature and Solid State Group of the National Research Council of Canada for harbouring me for some time during which this work was started.

APPENDIX

Solutions for the Phonon Equations of Motion in the Elastic Limit

The starting point is the determinantal equation given by deLaunay (1956), eqn. (10.9) page 266:

$$\begin{vmatrix} [(c_{11}-c_{44})q_x^2+c_{44}-\rho v^2] & (c_{12}+c_{44})q_xq_y & (c_{12}+c_{44})q_xq_z \\ (c_{12}+c_{44})q_xq_y & [(c_{11}-c_{44})q_y^2+c_{44}-\rho v^2] & (c_{12}+c_{44})q_yq_z \\ (c_{12}+c_{44})q_xq_z & (c_{12}+c_{44})q_zq_y & (c_{11}-c_{44})q_z^2+c_{44}-\rho v^2 \end{vmatrix} = 0, \quad \ldots \quad \text{(A 1)}$$

where we have let deLaunay's frequency ν be qv, and where $q_xq_yq_z$ are the direction cosines (not the components). If we make the ansätz

$$\rho v^2 = c_{44} + (c_{12}+c_{44})\lambda \quad \ldots \quad \text{(A 2)}$$

then (A 1) becomes with the definition (5.7)

$$\begin{vmatrix} sq_x^2-\lambda & q_xq_y & q_xq_z \\ q_xq_y & sq_y^2-\lambda & q_yq_z \\ q_xq_z & q_yq_z & sq_z^2-\lambda \end{vmatrix} = 0. \quad \ldots \quad \text{(A 3)}$$

This equation determines the direction cosines of vibration, which therefrom depend only on the direction cosines $q_x q_y q_z$ of propagation and on s_0. It was solved numerically. One could attempt a calculation using the solution of a Born–von Karman model, but we have not done it for the phonon-drag problem.

We now wish to prove (3.5) for this spectrum:

$$q \,.\, \nabla_q \omega = qv. \qquad \text{(A 4)}$$

Using (A 2), we get for our spectrum

$$\mathbf{q} \,.\, \nabla_q \omega = \mathbf{q} \,.\, \nabla(qv) = qv + q\frac{c_{12}+c_{44}}{2v\rho}\,[\mathbf{q} \,.\, \nabla\lambda]. \qquad \text{(A 5)}$$

To get $\mathbf{q} \,.\, \nabla\lambda$ we appeal to the following version of (A 3):

$$\lambda^3 - s\lambda^2 - (1-s^2)I_2\lambda - (1-s)^2(2+s)I_3 = 0, \qquad \text{(A 6)}$$

$$I_2 = q_x{}^2 q_y{}^2 + q_y{}^2 q_z{}^2 + q_z{}^2 q_x{}^2, \qquad \text{(A 7)}$$

$$I_3 = q_x{}^2 q_y{}^2 q_z{}^2. \qquad \text{(A 8)}$$

Then using the fact that $\mathbf{q} \,.\, \nabla I_2 = \mathbf{q} \,.\, \nabla I_3 = 0$, because

$$\mathbf{q} \,.\, \nabla q_i = 0. \qquad \text{(A 9)}$$

(remember that the q_i are the direction cosines), we easily obtain from (A 6)

$$\mathbf{q} \,.\, \nabla\lambda = 0. \qquad \text{(A 10)}$$

Hence (A 5) reduces to (A 4).

References

Bailyn, M., 1956, Thesis, Harvard University. Available on Microfilm only; 1958, *Phys. Rev.*, **112,** 1587; 1960, *Phys. Rev.*, **1,** 1960.

Bardeen, J., 1937, *Phys. Rev.*, **52,** 677.

De Launay, J., 1956, *Solid State Physics*, **2,** Ed. Seitz and Turnbull (New York: Academic Press Inc.).

Hanna, I. I., and Sondheimer, E. H., 1957, *Proc. roy. Soc.*, **239,** 247.

Huntington, H. B., 1958, *Solid State Physics*, **7,** Ed. Seitz and Turnbull (New York: Academic Press Inc.).

MacDonald, D. K. C., 1954, *Physica*, **20,** 996.

MacDonald, D. K. C., Pearson, W. B., and Templeton, I. M., 1958, *Proc. roy. Soc.* A, **248,** 107.

Mott, N. F., and Jones, H., 1936, *The Theory of the Properties of Metals and Alloys* (Oxford: Clarendon Press).

Neighbors, J. R., and Alers, G. A., 1958, *Phys. Rev.*, **111,** 707.

Parkinson, D. H., 1958, *Reports on Progress in Physics*, **21,** 226.

Roberts, L. M., 1957, *Proc. phys. Soc. Lond.*, **70,** 744.

Tsuji, M., 1959, *J. Phys. Soc. Japan*, **14,** 618.

Ziman, J. M., 1956, *Canad. J. Phys.*, **34,** 1256; 1959, *Phil. Mag.*, **39,** 371.

ERRATA

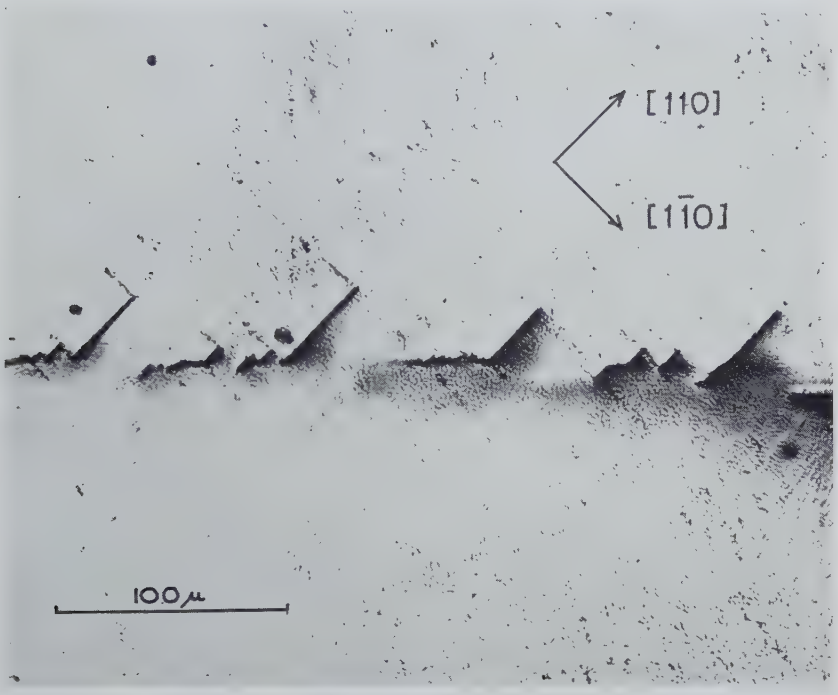
Palaeomagnetism of the British Carboniferous System, by C. W. F. Everitt and J. C. Belshé, 1960, *Phil. Mag.*, **5**, 675.

In the last column of the table on page 677 the ages of some of the rock formations are in error. The corrected figures are as follows.

Formation		Age (millions of years)
Staffordshire		240
South Wales		230–280
English Midlands	Sills	230
	Lava	270
Derbyshire	Sill	250 ?
	Lavas	270
Scotland (Kinghorn)	Lavas	280
	Sill	270 ?

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

Fig. 3

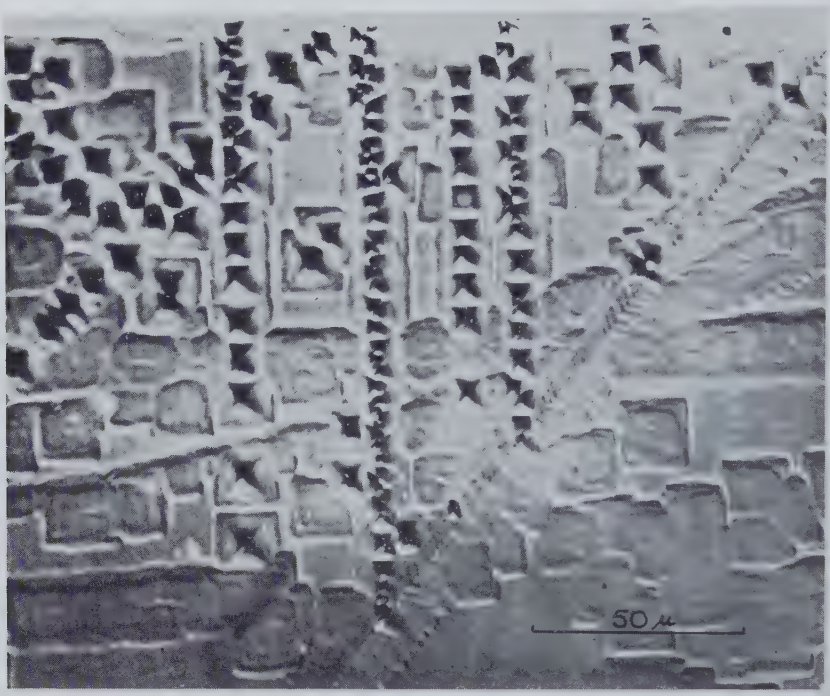


Stroh-type cracks in lithium fluoride.

Fig. 4

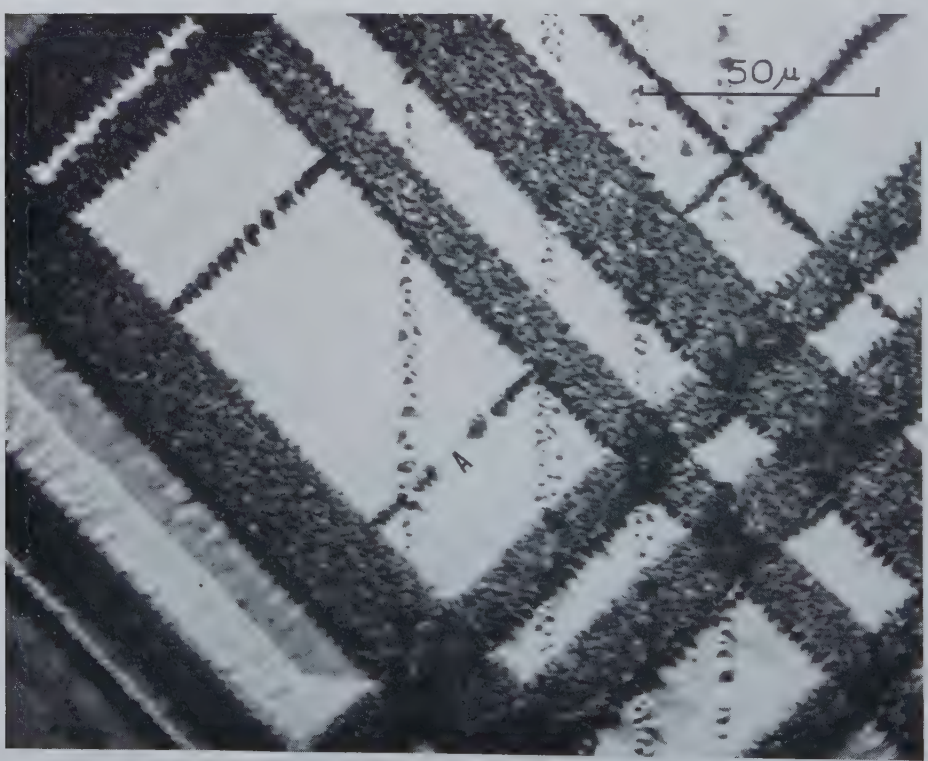
A pile up of edge-dislocations at a sub-boundary.

Fig. 5



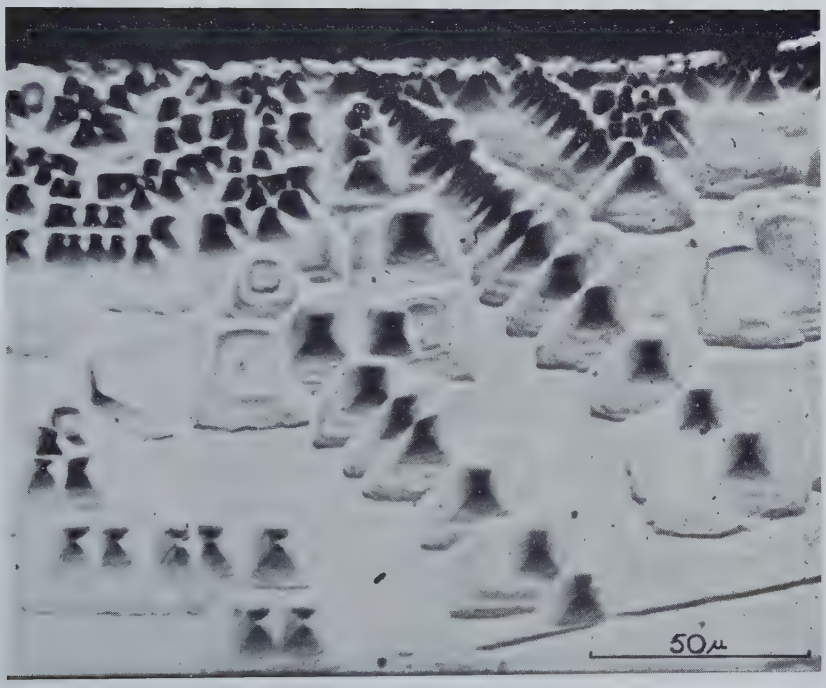
A pile up of screw-dislocations at a sub-boundary.

Fig. 6



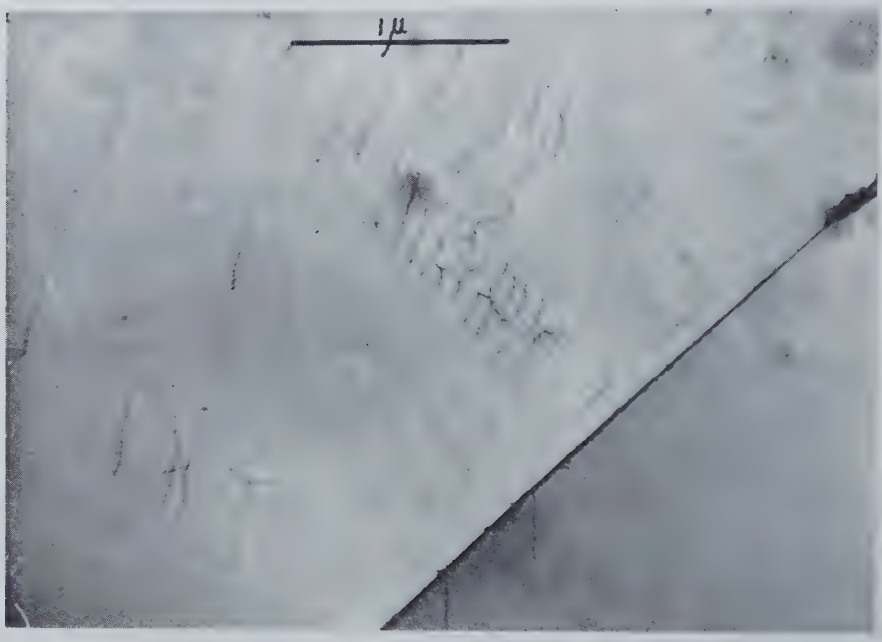
Broad-bands of edge dislocations acting as barrier to edge dislocations on intersecting system. Pile up of dislocations apparently from half-loop source at A.

Fig. 7



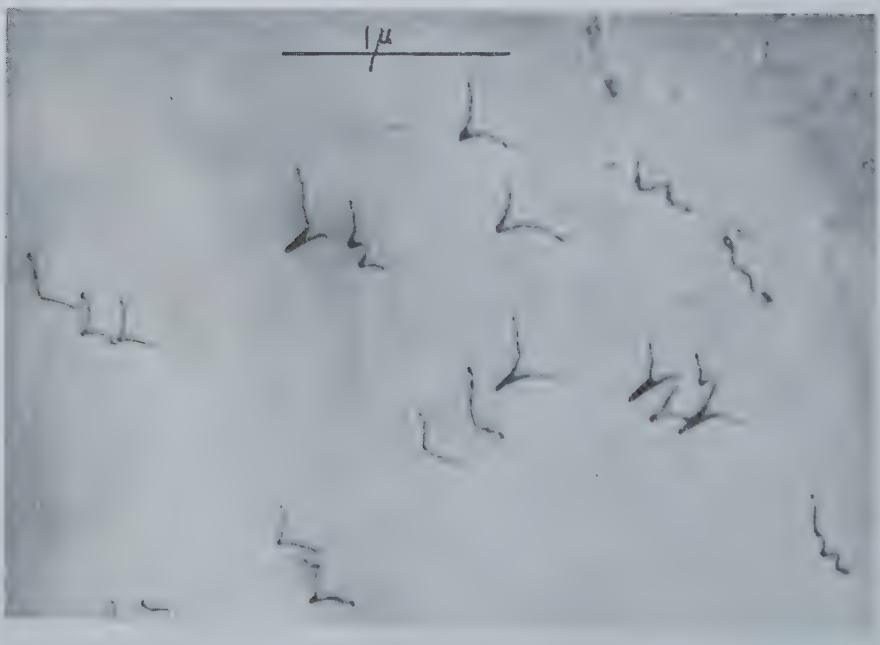
A pile up of edge dislocations beneath a surface coating.

Fig. 2



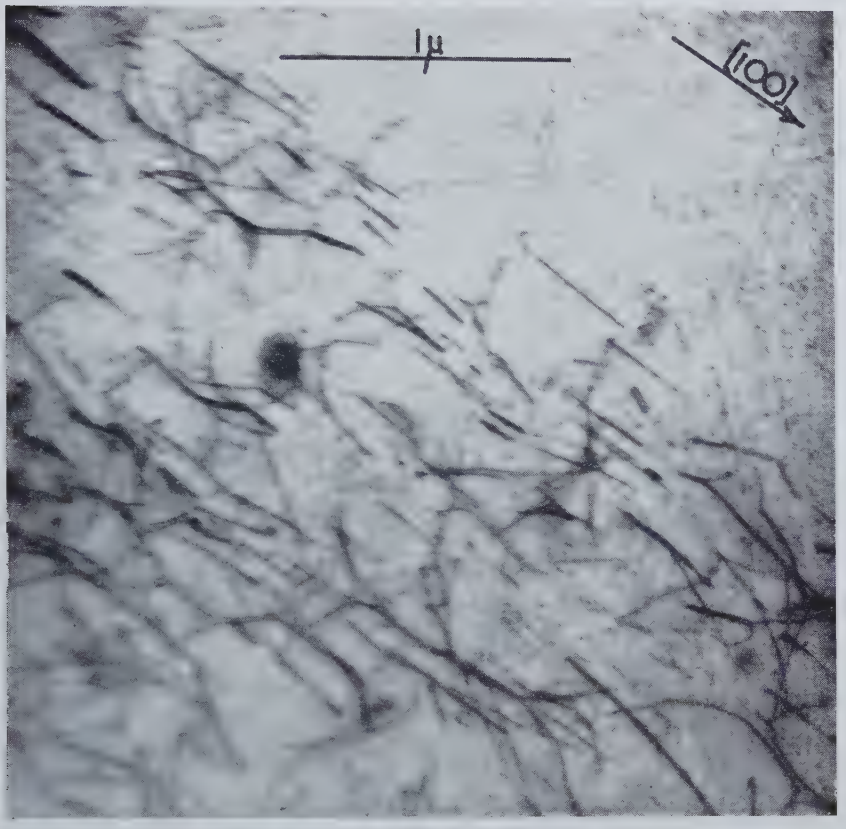
Screw dislocations on 45° planes. The diagonal line across the bottom right-hand corner is a cleavage crack.

Fig. 3



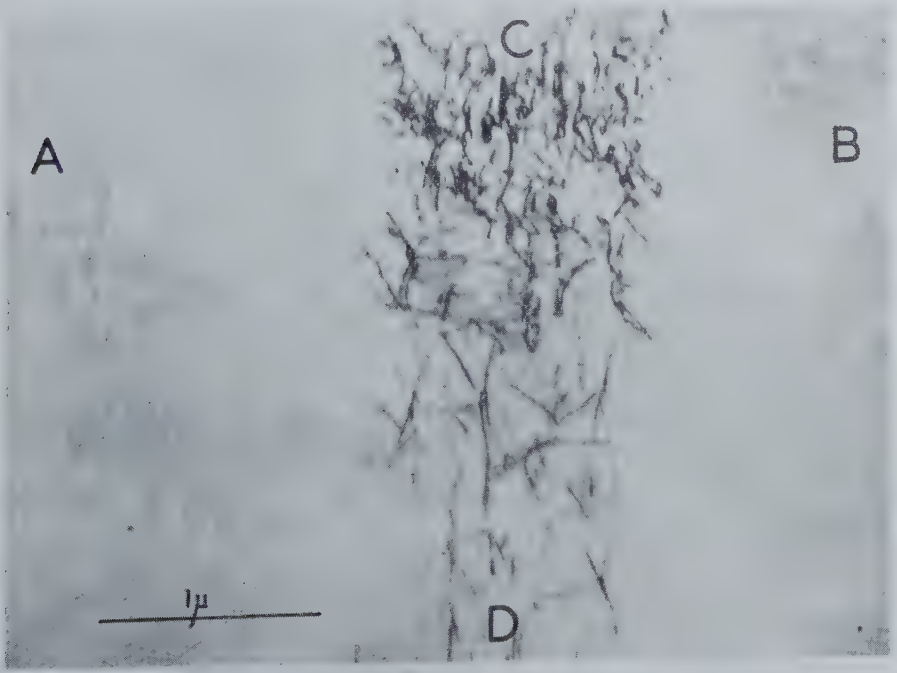
Dislocations on 45° planes after local heating in the electron beam.

Fig. 4



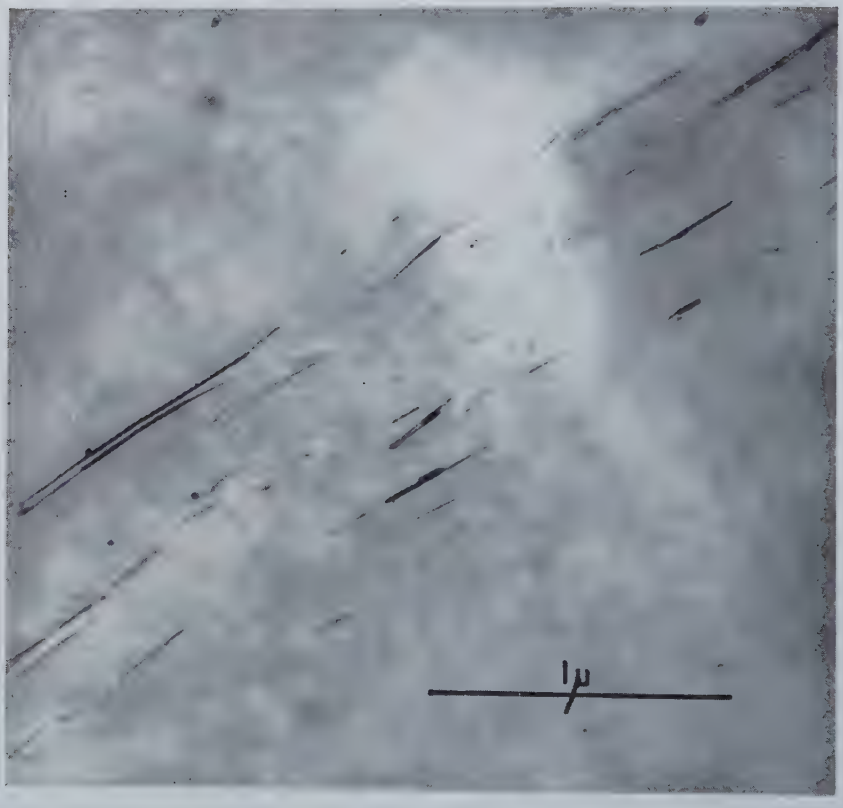
45° slip band showing plus–minus pairs.

Fig. 5



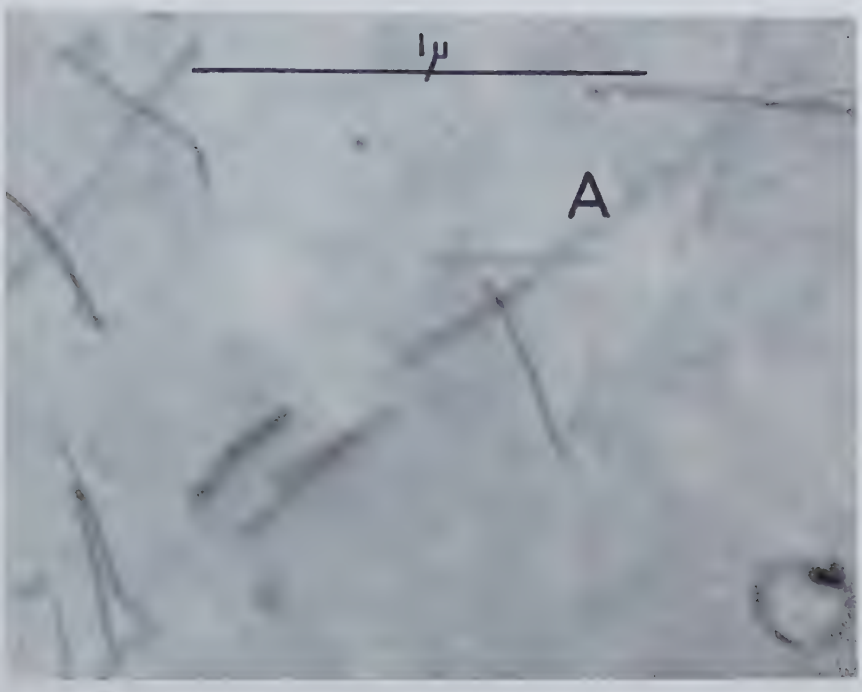
Two intersecting 45° slip bands. The band AB is almost out of contrast, though the tangle at the intersection of slip bands AB and CD is clearly visible.

Fig. 6



90° slip band.

Fig. 9

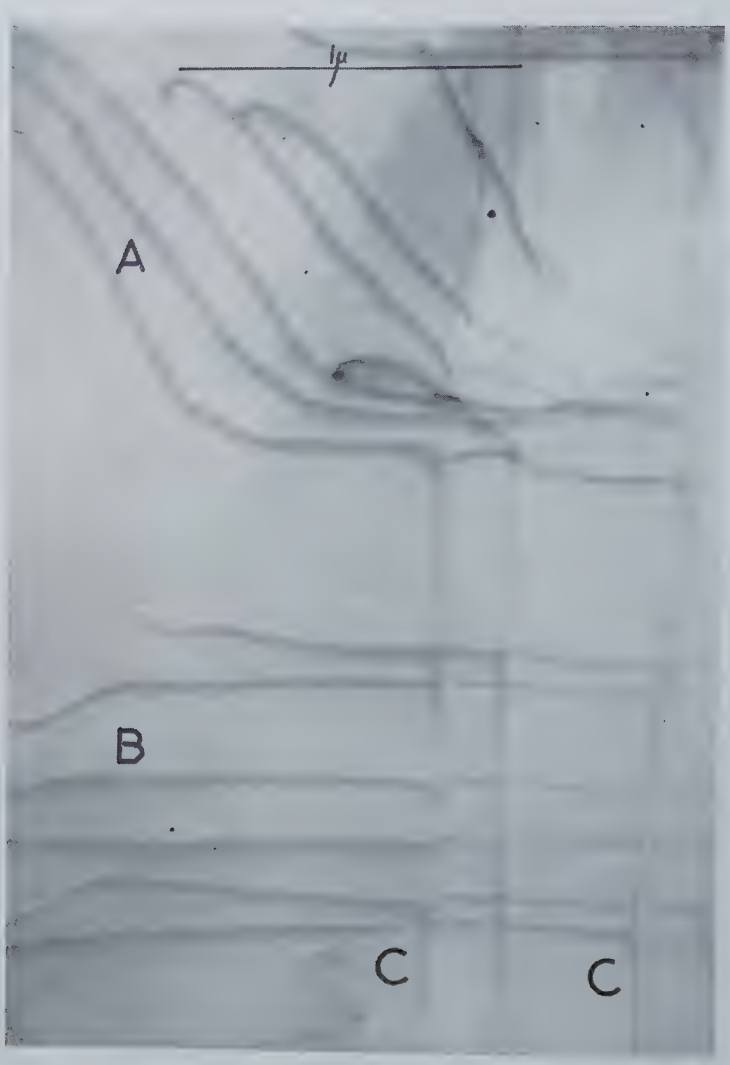


(*a*)

(*b*)

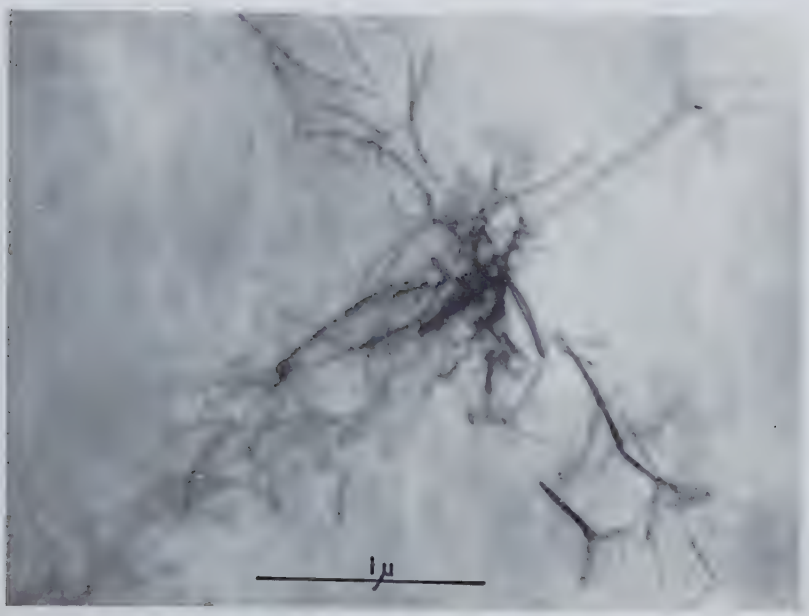
(*a*) Area containing plus–minus pairs before heating. (*b*) (Same area as *a*) after heating. The pair at A has broken up into loops.

Fig. 10



Dislocation interactions in MgO.

Fig. 11



Intersection of two 45° slip bands.

Fig. 12

Intersection of two 90° slip bands.

Fig. 2

Short surface crack terminated by slip at tips. Slip lines revealed by selective tarnishing. ×700.

Fig. 3

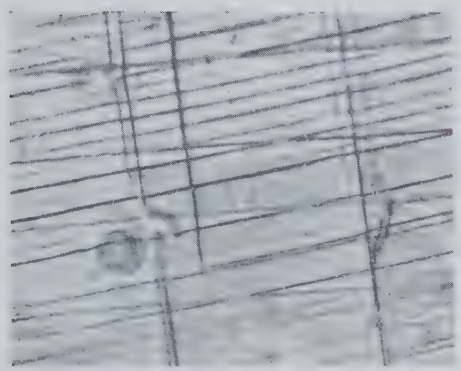
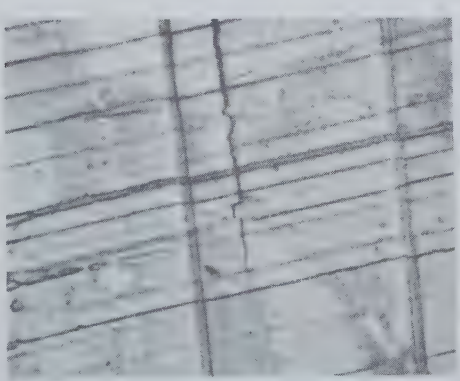


Fig. 4

Fig. 5

Fig. 6



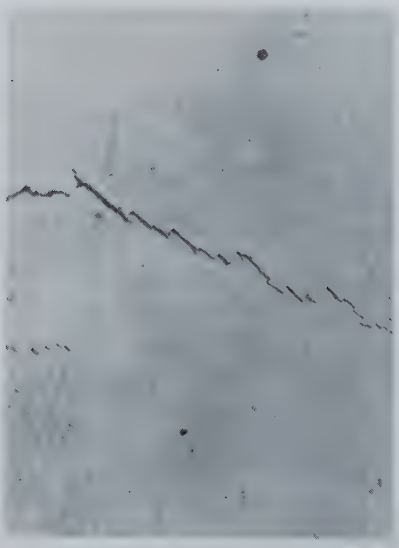
Figs. 3–6

Successive stages in the propagation of a stress-corrosion crack. ×1400.

Fig. 8

Crack pattern when operating slip plane is inclined to crack direction. ×450.

Fig. 9

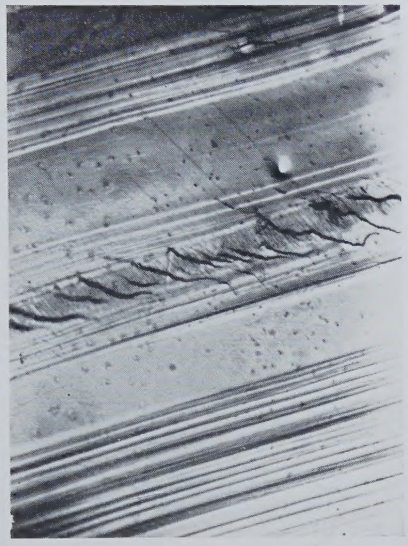


Transverse section through stress-corrosion cracks. ×500.

Fig. 10

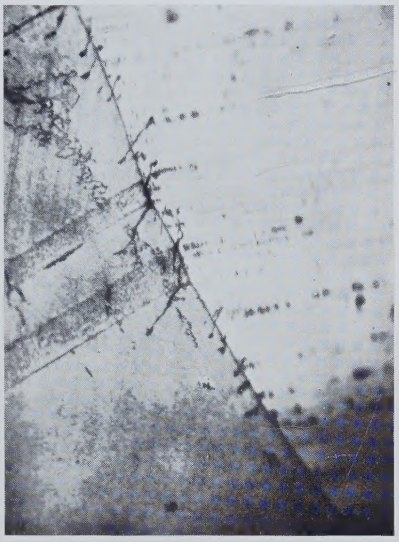
Formation of microcracks on the surface of a single grain at the interaction of slip bands. ×450.

Fig. 11



Cracks formed inside heavy slip bands. ×450.

Fig. 12



Formation of microcracks at grain boundary. ×450.

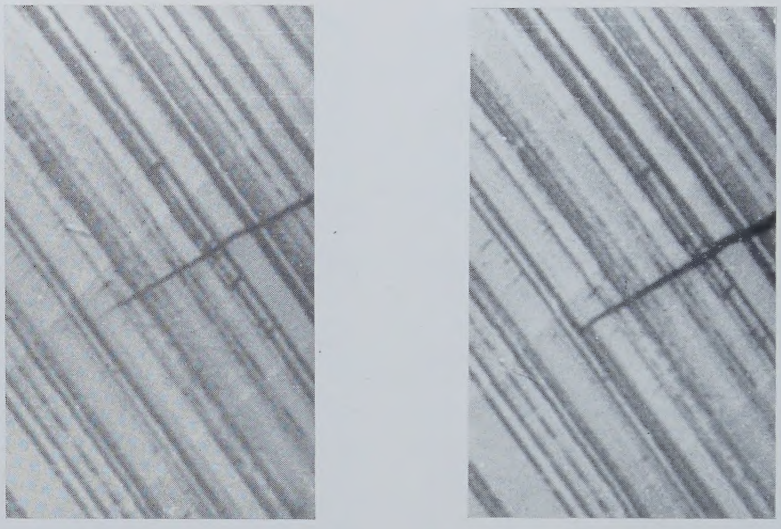
Fig. 13

Fig. 14

Figs. 13–14

Examples of extension of microcracks to produce grain boundary parting (principal tensile stress horizontal). ×450.

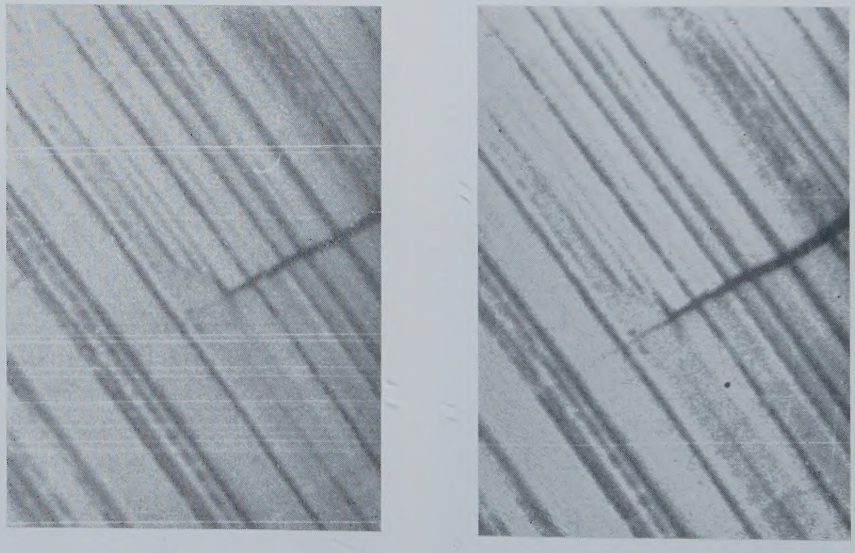
Fig. 15



(*a*) (*b*)

Showing the plastic yielding around the tip of a newly halted crack on increasing the applied load. ×2000.

Fig. 16



(*a*) (*b*)

Further cracking beyond the tip of an aged crack after increasing the applied load. ×2000.